WORKING WITHIN WIKIPEDIA: INFRASTRUCTURES OF KNOWING AND KNOWLEDGE PRODUCTION

A Thesis
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Master of Arts
in Communication, Culture, and Technology

By

Richard Stuart Geiger II, B.A.

Washington, DC
May 1, 2009
Acknowledgements

This thesis is dedicated my personal infrastructure of knowing and knowledge production – the vast assemblage of individuals who I have relied on for guidance and support throughout my time as a graduate student. These include my advisor David Ribes, who has helped me navigate the murky waters of academia as well as my reader Linda Garcia, who has continually provided new and enlightening perspectives. My fellow members in my thesis colloquium, Jed Brubaker and Margarita Rayzberg, have been critical in helping me work through my own ideas. In addition, there are countless individuals who have helped me stay sane throughout my time at Georgetown – an entire thesis could be written revealing such a powerful network of friends, colleagues, professors, and family.

R. Stuart Geiger
WORKING WITHIN WIKIPEDIA:
INFRASTRUCTURES OF KNOWING AND KNOWLEDGE PRODUCTION

R. Stuart Geiger II, B.A.

Thesis Advisor: David Ribes, Ph.D.

ABSTRACT

Wikipedia, the self-proclaimed “free encyclopedia that anyone can edit,” is often mystified in the public and academic imagination. Because Wikipedia allows anyone to edit any article at any time in any manner without review, common sense is astounded when it is claimed that the encyclopedia exhibits roughly the same level of quality as the Encyclopedia Britannica. This is largely because popular representations of Wikipedia hold that the project's software system, social structure, lack of management, and cultural values all seem to be built precisely to discourage such regulation. As such, the fact that Wikipedia works remains a mystery, seemingly contradicting everything that we as a society know about how humans organize in groups. In response, researchers from many disciplines have attempted to forge explanations of the encyclopedia project’s order and regularity by adapting existing theories to Wikipedia. More sociologically-inclined researchers argue that Wikipedia is more like a traditional society, held together by shared norms, rules, traditions, roles, institutions, discourses, and other sociocultural macro-structures. Another tradition has compared Wikipedia to a market-based economy in which editors interact in unmediated and uncoordinated
micro-level ‘transactions.’ In this thesis, I critique both of these depictions of Wikipedia, which take for granted the project’s largely-invisible and unofficial infrastructures of knowing and knowledge production.

While the ‘anyone can edit’ functionality of Wikipedia’s software architecture is essential, social and economic explanations of Wikipedia have generally failed to take into account technology in Wikipedia beyond this generic feature. As I show, there are a significant number of transformative technologies which have been deployed both officially and unofficially. Order in Wikipedia emerges not merely from social or economic forces, but alongside complex assemblages of technical and social actors who work to associate, aggregate, delegate, routinize, contextualize, and operationalize the seemingly-autonomous act of contributing to a Wikipedia article. I argue that research into Wikipedia must take into account the diverse and heterogeneous assemblage of humans and non-humans who are constantly working to make Wikipedia into the kind of place where either social or economic forces can come on the scene to explain how millions of individuals can work together to produce the largest encyclopedia in human history.
Table of Contents

Chapter 1. Introduction: “Bots Are Editors, Too!” ................................................................. 1

Chapter 2. Literature Review .............................................................................................. 18
   Introduction .................................................................................................................. 18
   Scholarly Studies of Wikipedia .................................................................................... 23
   The Social Construction of Wikiality .......................................................................... 26
   The Rise and Fall of The Wisdom of Crowds .............................................................. 38
   The Normative Structure of Post-Production Control .................................................. 56

Chapter 3. Bots .................................................................................................................. 71
   Introduction ................................................................................................................ 71
   The Case of HagermanBot ......................................................................................... 75
      A Task and a Request ............................................................................................. 75
      Responding to Objections ...................................................................................... 79
      An Unexpected Ally ................................................................................................ 87
      Blackboxing the Compromise ................................................................................ 91
   Realigning the Network: The Case of Betacommand’s Bots ........................................ 97
   Analysis ....................................................................................................................... 106
      The Big Leviathan ................................................................................................... 106
      Actualizing Consensus ............................................................................................ 113
      Circulating Blackboxes .......................................................................................... 117
   Conclusion .................................................................................................................. 121

Chapter 4. Vandal Fighting .................................................................................................. 125
   Introduction ................................................................................................................ 125
   Problematizing Macro-Actors ..................................................................................... 133
   Vandal Fighting: The Case of 72.68.228.176 ............................................................. 141
      Technologies of Knowing ....................................................................................... 141
      A Vandal Emerges ................................................................................................. 144
   Administrator Intervention against Vandalism (AIV) .................................................. 154
   Analysis ....................................................................................................................... 162
   AIV Statistics .............................................................................................................. 166
   Conclusion .................................................................................................................. 173

Chapter 5. Conclusion ........................................................................................................ 175

Chapter 6. Bibliography ...................................................................................................... 192
Introduction: “Bots Are Editors, Too!”

In late 2006, editors in the English-language version of Wikipedia began preparing for their third annual election for members of the project’s Arbitration Committee. In its own words, the Committee “exists to impose binding solutions to Wikipedia disputes that neither communal discussion, administrators, nor mediation have been able to resolve.” ArbCom (as it is known colloquially) is often compared to a high court, and membership is one of the most formal and powerful positions of authority; as such, elections for the few seats that open each year are hotly contested by candidates and voters alike. In 2006, nominations for open seats were accepted during the month of November, and according to the established rules, all Wikipedians who made at least 1,000 edits to the encyclopedia project as of October of that year were eligible to run.

In all, about forty editors meeting these requirements nominated themselves or accepted the nominations of others, which formally involved submitting a brief statement to potential voters with reasons why they would be good arbitrators. One such candidate was a user named AntiVandalBot, a computerized program that reviewed all edits to the project as they were made and reverted those that, according to its algorithms, were blatant acts of vandalism. This ‘bot’ was written and operated by a user named Tawker, who used the separate account to distinguish edits he made with those made by the program – a well-established convention. In AntiVandalBot’s statement to potential voters, the computerized editor described its accomplishments and adherence to project policies (like Neutral Point of View, or NPOV) in the same manner as many other
candidates, even replying on common rhetorical tropes:

In my 6 months+ of protecting Wikipedia from vandalism attacks I have never once made an NPOV edit, something that is very hard to accomplish with human editors. I always express NPOV on any decision I make because I have no intelligence, I am only lines of code. I also never tire, I work 24 hours a day, 7 days a week. I think I have the most of edits of any account on this Wiki now, I have not counted since the toolserver database died. Taking a look at my talk page history, my overseers ensure that all concerns are promptly responded to. In short, a bot like me who can function as a Magic 8 Ball is exactly what we need on ArbCom! -- AntiVandalBot 05:20, 17 November 2006 (UTC)

Once a nomination has been put forth, voters regularly submit public questions to specific candidates, and AntiVandalBot was not exempt from this practice. One user asked the bot its position on Isaac Asimov’s laws of robotics and their applicability to Wikipedia. Others asked questions which implied that the bot was not a well-rounded editor, claiming it was too involved in fighting vandals and not involved enough in other areas, like dispute resolution – a common criticism leveled against many such candidates. While these Wikipedians treated the bot with at least an ironic level of seriousness, others were offended at the bot’s emergence in the election. One user removed the bot’s candidate statement from the election page without prior discussion; however, AntiVandalBot reverted this deletion of content, identifying it as vandalism in the same way it did with hundreds of thousands of edits before. An edit war sparked over the issue, as Wikipedians with various levels of seriousness removed the statement, argued that the bot had no place in the official election, and criticized Tawker for nominating his bot. In response, Tawker repeatedly replaced the bot’s statement on the nomination page and passionately defended the right of bots to run for election if they had such an inclination of their own free will – and he claimed that his did. When a user asked
AntiVandalBot if the repeated removal and replacement of its candidate statement on the nomination page “indicate[s] that you don't really want us to take you seriously,” the bot staunchly defended itself in this political public sphere by exclaiming, “Bots are editors too!”

I make the same argument in this thesis, although in a markedly different context. Tawker – speaking through his bot – was jokingly claiming that computerized editors ought to have the same sociopolitical rights and responsibilities as human editors, capable of running for the project’s highest elected position and, if successful, influencing the process of encyclopedia-building at its most supreme level. In contrast, I argue (with all seriousness) that these software agents already have a similar level of influence on how Wikipedia as an free and open encyclopedia project is constituted, fundamentally shaping the way in which all users – administrators, power users, casual contributors, and even vandals – exist as editors. However, previous academic research and popular discourse about Wikipedia has generally passed over these technological actors, giving explanations of the encyclopedia project that are almost exclusively based in purely social structures and artifacts. While there is no doubt that the project’s shared norms, codified standards, administrative processes, and formal institutions like ArbCom work hard to hold together the social order through which encyclopedia articles are produced and negotiated, I argue that such purely social actors do not exhaust explanations of Wikipedia. Rather, the encyclopedia project’s unlikely and unexpected success must also be attributed to the work of automated, computerized editors like AntiVandalBot, as well as a whole host of technological actors. Such non-human actors collectively
but contingently comprise an infrastructure of knowing through which Wikipedians – knowingly or not – come to know their project, its content, and those who edit it.

One of the most important aspects of these software agents is that most are unofficial, that is, not directly incorporated into the project’s software architecture – known as MediaWiki. If a systems administrator downloads this free, open-source software and installs it on a web server (as hundreds of thousands have done, including myself), the empty wiki site will look and feel almost exactly the same as Wikipedia, but only on the surface. As I personally discovered, the MediaWiki-based site I established – which was to help high school and college policy debaters collaboratively craft arguments – was inundated not with the wisdom of the crowds but the guile of spammers and vandals. In order to keep the wiki usable in any form, I had to lock down access to such an extent that it has became a mausoleum. While I hesitate to attribute the failure of this project merely to vandals and spammers, I have since learned that my experience was not unique. I have met dozens of other wiki administrators, most of whom had far more promising projects than mine, with similar experiences and the same level of confusion. Everything that common sense, sociological theory, and critics of Web 2.0 predicted regarding Wikipedia and its open editorial model had come true, but only on our wikis.

Such failures are often explained away by one of the most common rationalizations of the wiki model: the constitution of the project’s editorial base, particularly its size. Under this view, my small, open wiki worked perfectly to capture the collective ‘wisdom’ of the ‘crowd’ who edited it: about ten percent wanted the wiki to be a shared resource for crafting arguments, another ten percent wanted a
virtual graffiti wall, and the remaining eighty percent wanted a space to advertise pornography and discount pharmaceuticals. If my wiki had a dedicated editorial base even a fraction of the size of Wikipedia’s, there would be enough users working on various aspects of the project that vandalism and spam would be reversed rather quickly.

In this kind of explanation, the point of failure for my site was nothing inherent in the wiki software, the wiki model, or even my specific project, but rather a function of the kinds of people who had freely decided that it was in their best interests to edit the site in a certain way. If I wanted to be more like Wikipedia – the project that nearly all wiki-based projects are inspired by – I would have to figure out a way to get enough good contributors interested in this site that they would balance out the vandals and spammers that inevitably flock to any public wiki.

However, it seems counterintuitive that the increased size of a group would lead to more order, not less, even if the members held near-identical worldviews – which is most certainly not the case. Such an explanation tends to invoke violent images of mob rule among critics, a social order that seems better suited for burning encyclopedias, not writing them. Yet Wikipedia not only persists, but thrives: it is the largest encyclopedia in human history and, as popular and academic accounts have claimed, it is far more accurate than one would expect. This phenomenon is especially difficult to reconcile given that its members include those who can enter Wikipedia’s largely-anonymous crowd without knowing anything about the project or the article they wish to change. In response, a growing number of social network theorists, most notably James Surowicki, Yochai Benkler, and Cass Sunstein, have argued against traditional sociological and
organizational explanations of group behavior in these “networked economies.” They claim that sites like Wikipedia are part of a new phenomenon in contemporary society in which social groups are better conceptualized using a model of decentralized social networks instead of hierarchical social structures. Taking largely from economic theories of rational choice, these theorists argue that “the wisdom of crowds” or “the wealth of networks” is an emergent phenomenon made possible by the staggeringly low costs involved with interacting with any other user in such a collaborative space. In the same way that a free market chooses the ideal price of a good through the myriad of uncoordinated interactions between freely-associated individualistic agents, so do these social networks choose the ideal presentation of knowledge in Wikipedia. By simply giving all users the ability to make any change to any article in any manner, limited only by the fact that all other users have this same ability, sites like Wikipedia make possible a new form of collective intelligence that operates from the bottom-up, instead of top-down. Given a crowd or network of the right size and temperament (and the theorists differ on the specifics) and a mechanism or medium for facilitating micro-level interactions between all members, the emergence of order is not an anomaly, but to be expected.

My issue with this explanation is that it only accounts for the project’s software architecture insofar as the technology enables connections between any two individuals in the same way that a market’s infrastructure enables interactions between buyers and sellers. In Wikipedia, this is re-framed as the ability for any editor to contest any other edit, removing it if a particular editor feels that it violates their view of what an
encyclopedia article ought to look like. However, as I show in this thesis, the software extends far beyond this simple ability which is present in any wiki-based website. Behind the scenes, contributions made are being endlessly transformed, circulated, aggregated, normalized, operationalized, and distributed by largely-unofficial software agents to such an extent that it makes no sense to describe editors in Wikipedia as disconnected, uncoordinated individuals. By the time one Wikipedian editor – from the most elite power user to the most novice anonymous vandal – ‘seamlessly’ makes a contribution in relation to another’s edits, a multitude of algorithmic actors have already arrived on the scene to facilitate such a process in a socially-amenable manner. While editing Wikipedia may seem like an autonomous process in which self-determining humans are able to independently act without needing to coordinate with each other, this is not an \textit{a priori} phenomenon of networks, but rather of the software agents who must constantly be deployed in order to stitch the social fabric together out of these fragmentations. In short, the empirical claims that social network theorists make regarding the emergence of order in Wikipedia are not necessarily false, but rather only true because of a staggering amount of work performed by human and non-human actors allowing humans to contribute as if they were disconnected, independent agents.

Software agents have not only been neglected by social network theorists, but most social scientific scholars as well. While many researchers have studied Wikipedia’s social structure – including its norms, ideals, standards, dispute resolution mechanisms, social roles, governing organizations, and so on – bots and other software agents have been passed over, or worse, explicitly ignored. Like the social network theorists,
most researchers who attempt to explain Wikipedia’s order posit a predominantly a-technical force that works behind the scenes to structure the way in which contributors to Wikipedia act in coordination – in spite of themselves. The largely social artifacts, norms, roles, and structures may be only possible because of the technological medium, but like the network theorist view, such an analysis does not extend beyond the “anyone can edit” feature provided by the software. Like in the tradition of social network theory, technology has not been a key primary variable in explaining why Wikipedia of all wikis exhibits some semblance of order, because this functionality is present in all wikis. While network theorists identify micro-level transactions between a sufficiently large and well-constituted crowd as the force behind Wikipedia’s order, this avenue of research has uncovered overarching sociocultural structures that regularly articulate and enforce a shared vision of how the free and open encyclopedia project is and ought to be. If the goal of the social network theorists is to show that an organization like Wikipedia is a fundamentally new kind of entity that cannot be explained using traditional sociological and organizational theories, then these researchers can be seen as the expected rebuttal from these disciplines. In such research, Wikipedia is made far more familiar to off-line forms of social organization, and some draw explicit comparisons to groups who are organized around commonly-held natural resources. Conclusions range from the celebratory to critical, but tend to reject the spontaneously emergent, micro-level depiction of Wikipedia posited by social network theorists in favor of one that relies more on norms, institutions, social hierarchies, discursive orders, and governance mechanisms to show how the project can organize against the barbarians at the gates – vandals,
ideologues, spammers. However, this approach tends to problematically posit the existence of overarching social structures to explain the way in which editors act in Wikipedia, in spite of themselves. The observed order within Wikipedia is generally attributed to the work that these macro-actors – such as governance, policy, discourse, and culture – perform on individual editors.

While sociologically-themed research of Wikipedia has revealed many important aspects about the project, it has consistently passed over most of Wikipedia’s technological infrastructure. I argue that this leads such research to attribute a rationality to Wikipedia’s social structures while ignoring the active roles of such software agents in constituting such macro-actors. The most striking example of this are bots, the automated editors who unceasingly work to maintain various aspects of the project. These software agents are systematically excluded from social scientific research, particularly in quantitative research generated from logging data, and often without a rationale. Researchers who do give explanations for rejecting bots in their analysis generally cite findings made from data collected in 2005, which found that bots comprise only 3 to 5 percent of all edits. At the time, bots were also generally involved in single use tasks which would have been performed by human editors but was simply too tedious – capitalizing every instance of “united states of america,” for example. While these kinds of bots still exist, new kinds of bots have come on the scene, and I present new evidence that demonstrates the explosive growth in bots as of 2009. Some of the most active bots those like AntiVandalBot, which patrol for vandalism and automatically revert blatant incidents. Others work to maintain order in discussion spaces,
automatically adding signatures to unsigned comments made on discussion pages and archiving old conversations. Another group of bots perform administrative maintenance work, sorting, classifying, organizing, and delegating administrative requests to block users or temporarily protect pages from editing. Bots also deliver newsletters to interested users, collect statistics, flag articles which meet (or fail to meet) certain criteria, and more. In all, over nine hundred bots have been officially approved since 2002, although there are only about one hundred bots currently in operation as of 2009.

Not only have bots grown significantly in use, a new kind of technological tool has emerged on the scene: assisted editing programs. The traditional method of editing encyclopedia articles is to review them, click the “edit this page” button, make whatever changes are deemed necessary, and then submit; assisted editing programs speed up certain kinds of editing work by significantly by automating various elements of the process. One class of programs allows users to view all edits made to Wikipedia in real time, and for the sake of convenience many customized filters are often used. For example, a user can choose to review only those edits which have commonly misspelled words, telltale signs of vandalism, or those made by anonymous users – among many other criteria. Another class of assisted editing programs automates routine actions that are not directly involved with encyclopedia editing, but still essential to the well-functioning of the project, such as reporting disruptive users to administrators.

In order to determine the relative prevalence of these software agents, I first recorded of all edits was made to Wikipedia using a bot developed under the SxWiki PHP framework. This bot, which made no edits and was operated anonymously,
gathered a list of all recent changes made to the project and ran on an hourly basis for two months. In this way, data was collected for every edit made to Wikipedia between 4 February 2009 and 1 April 2009 – 12,352,612 edits in total. I then coded the edits by username to determine bot status, taking from various lists of bots compiled by bot operators. Assisted editing programs were detected by the presence of various indicators inserted by such programs. The results are striking, especially compared to findings made in 2005 and 2006 that bots only make 3 to 5 percent of edits. In the observed period, bots comprised 16.33% of all edits, while assisted editing programs were used in approximately 12.16% of all edits. Graphing the edits made in each user category each day (Figure 1) shows that assisted editing is relatively constant, while bot operation is more variable. The breakdown of assisted edits by program (Figure 2) reveals many different tools in active use, with a few comprising the vast majority of assisted edits.

Furthermore, an analysis of log data provided by the Wikimedia Foundation indicates that the growth of bots has increased dramatically since 2005. Similar analysis could not be conducted for assisted edits, as the log data released by the Wikimedia Foundation does not include data necessary to determine which edits were made using assisted editing tools. A graph of edits per month from 2003 to 2008, stacked by user type (Figure 3) shows that the number of bot edits has steadily increased over Wikipedia’s history. This data also confirms the findings of researchers made in 2005 regarding the miniscule roles of bots at that time, indicating that the proportion of edits has indeed grown substantially. A graph stacked proportionally (Figure 4) shows that since early 2007, bots have made more edits to Wikipedia than administrators.
Figure 1.1: Daily edits per user group, stacked proportionally
Figure 1.2: Daily edits for assisted edits, stacked proportionally by tool
Figure 1.3: Monthly edits by user group, stacked
Figure 1.4: Monthly edits by user group, stacked proportionally
In all, these findings indicate that automated and assisted software agents have a much greater role in Wikipedia than previous research has suggested. The issue at hand is the extent to which these technological actors are active participants in the Wikipedian lifeworld, and as such, how their introduction has transformed the way in which Wikipedia operates. In order to answer this question, I perform a highly empirical analysis of these tools as they are used by Wikipedian editors in their average, everyday activities. While many popular and academic accounts of Wikipedia claim that either social structures or a ‘wisdom of crowds’ invisibly guides its participants, this thesis paints a very different picture of the encyclopedia project’s order and stability.

In the first chapter, I take the reader through a history of ideas regarding Wikipedia, describing the way in which the encyclopedia project has been represented in the popular press and various academic disciplines. I identify and subsequently critique the predominant views of Wikipedia, showing how they have explained the project’s observed order in ways that take for granted the stability of society and do not take into account its technological infrastructure. While I do not explicitly reject the findings of research from either social network theory or sociologically-themed disciplines, I do argue that they have systematically ignored the role of technology in working behind the scenes to ensure that their vision of the world operates ‘seamlessly.’ In the second chapter, I give an account of two bots which were introduced into the community to realize a particular view of how their operators thought Wikipedia ought to operate. By tracing out the controversies that emerged in response to the bot’s introduction, I describe the kinds of normative activities that these actors make possible within Wikipedia.
also show how these bots themselves are made possible through a complex sociotechnical infrastructure that emerges out of neither a pre-existing social order nor an emergent phenomenon of networks, but is rather one that must be constantly and actively maintained.

In the third chapter, I explore the way in which collective intelligence is produced in Wikipedia by analyzing ‘vandal fighting,’ the process in which editors review contributions for malevolent or disruptive edits. I describe the way in which a vast assemblage of technological actors actively reshape the way in which editors experience their project and edits to it, making possible a kind of epistemological enforcement that requires little to no specific knowledge about the article at hand. While the individual humans described in this chapter seem to be disconnected agents who independently act in ways they see fit, order emerges as complex assemblages of technical and social actors work to associate, aggregate, delegate, routinize, contextualize, and operationalize the seemingly-autonomous act of contributing to an article.
Chapter 1. Literature Review

Introduction

Wikipedia, the self-proclaimed “free encyclopedia that anyone can edit,” is frequently depicted as an anarchic battleground of truth. For good or ill, millions of seemingly-disconnected individuals have somehow created the world’s largest and most popular reference work, and popular accounts often accredit the project’s success to its radical rejection of traditional notions of editorial authority and social organization. Yet an increasing number social scientists studying Wikipedia have concluded that the project is not anarchic, but instead held together by a social structure that is similar to that of many offline groups, particularly those organized around common economic resources. When individuals enter Wikipedia, these scholars claim, they enter a realm that many seem disorderly, but is in fact ruled according to a set of social forces that are largely invisible – that is, until one brushes up against them. As many critiques of Wikipedia have argued, these social forces are often strikingly different from those governing more ‘traditional’ societies of knowledge production, like science or journalism.

As such, there has been a shift in popular representations of Wikipedia away from the predominant view that the project is a primal ur-society whose members are shockingly adept at generating something resembling an encyclopedia. Around late 2006 and early 2007, Wikipedia became increasingly portrayed as a society with its own culture, rules, norms, and roles, which were of growing interest to many different kinds of individuals with a voice in the public sphere. Responding to the increasing number of
double-blind studies which generally showed that Wikipedia was about as accurate as other, more traditional encyclopedias, discussion of Wikipedia’s social structure turned into critiques which attempted to expose cases in which the social structure went awry. Such accounts, I argue, are reminiscent of the ‘Science Wars’ of the 1990s, as they advance their arguments by portraying knowledge in Wikipedia as a social construction, tarnished by cultural forces to the extent that its articles have no more than a tangential connection to truth.

Yet despite the increased frequency of such critiques, the fact remained that knowledge presented in Wikipedia was far more ordered than expected, leading many to reject sociological explanations of Wikipedia’s accuracy in favor of economics. Adherents of this view largely follow James Surowicki’s work *The Wisdom of Crowds*, using neoliberal economic theory to explain Wikipedia’s unexpected order. Surowicki uses mathematical theories of rational choice to formalize the conditions under which crowds can make wise decisions in aggregate. Specifically, he claims that attempts to externally coordinate behavior in a group are inefficient and lead to worse decisions, a claim that mirrors neoliberal views of government regulation. Instead, he claims that the best decisions are made by the raw economic aggregation of individual opinion in a marketplace of ideas, the ideal vision of this being a stock or betting market in which shares of truths and predictions can be bought and sold. This model has been applied to Wikipedia by many of Surowicki’s followers, who claim that the encyclopedia project chooses the most accurate facts in the same way that a market arrives at an ideal price. In contrast to the sociological view, this conception of Wikipedia holds that the
project’s disconnected and primal anarchism is what explains its congruence with the truth. However, an examination of Wikipedia shows that it does not function in this manner, primarily because the opinions of its crowd are aggregated through highly discursive and social processes of negotiation and influence. Other adherents in this view include Cass Sunstein, whose work *Infotopia* makes a similar argument and, unlike Suroweicki, specifically applies it to Wikipedia. While the breadth of scholarship in social network theory and neoliberal economics is far more complex than the narrative I tell, it has largely not been directly applied to Wikipedia outside of the spontaneously-emergent ‘wisdom of crowds’ types of explanations. My aim in this section is not to critique social network theory and neoliberal economics, but rather the way in which these theories have been problematically applied to account for Wikipedia’s unexpected success in both popular and academic spheres.

Next, I review a more sociologically-themed view of Wikipedia, advanced by researchers who assume that Wikipedia’s order can be explained by a highly-active group of ‘elites’ who construct norms and then enforce them among ordinary users. Various social artifacts, such as codified rules and templates, are taken to make this process easier by reducing coordination costs. For example, due to the pre-existing structure of various social orders like the project’s “policy environment,” interactions between users in the administrative class are mediated such that they can easily create new norms or punish users who violate existing ones. Yet despite such social structures, this work seems highly inefficient, especially to those who still hold a strictly-atomistic, rational choice view of Wikipedia, as it relies on a sizable group of trusted users to continually patrol
for disruptions and forcefully re-establish order. Because of this, many critiques have been made that this system is unsustainable due to its inability to retain existing enforcers and recruit new replacements.

As I show, even through these critiques have been empirically denied, such sociologically-heavy accounts are problematic because they largely ignore the specific effects of technological actors. By focusing almost exclusively on the reduction of coordination costs in social interaction, these studies tend to ignore the unique role play that various programs, tools, visualization mechanisms, and database systems play in holding together the Wikipedian social order. Those researchers who do examine technological tools in Wikipedia have either made speculative claims without empirical grounding, have not studied tools which are actually in use by the community, or have examined tools which have been present in the project since its founding. What is needed is a study of Wikipedia that examines both the way in which technological tools are introduced into Wikipedia as well as the way in which they materially transform the decision-making or aggregation process.

While the spontaneously-emergent ‘wisdom of crowds’ rebuttal to the social constructivist view of Wikipedia is untenable, the traditional sociological characterization of Wikipedia is not able to adequately contest such a ‘marvelous’ explanation of the project’s order and stability. In its place, this work will give an account of Wikipedia that accounts for the social and technical infrastructures of knowing and knowledge production through which the project is held together. Instead of taking society and technology as pre-existing entities which structure and explain human behavior, this
study shows the staggering amount of work which is regularly performed by humans and non-humans in order to keep these infrastructures in working order. Without falling back into an economic/mathematical incentives-based framework, this study explains the specific manner in which the Wikipedian society – a diverse and heterogenous assemblage of editors, administrators, articles, algorithms, and more – routinely produce order from disorder at great cost and to great reward.
Scholarly Studies of Wikipedia

Virtually all disciplines in the social sciences and humanities have been used to examine aspects of Wikipedia. Much has been written on the economics of article production, asking questions about labor, value, and the nature of public goods in the project (Polborn 2007; Ortega, Gonzalez-Barahona, and Robles 2008; Anthony, Williamson, and Sean Smith 2005). Wikipedia’s bureaucratic governance model has been compared to that of various political institutions around the world (Navot 2008; Noveck 2008). The motivations and personalities of Wikipedian contributors have been surveyed (Oreg and Nov 2008; Nov 2007; Kuznetsov 2006). The discourse of the Wikipedian community has been analyzed from various theoretical perspectives, both inside and out of communication studies (Pentzold and Seidenglanz 2006; Geiger 2008; Hansen, Berente, and Lyytinen 2007). Wikipedia's model of article writing has been compared with the process of historical writing (Rosenzweig 2006). Philosophers have examined the epistemological model present in Wikipedia, comparing it to various philosophical positions (P.D. Magnus 2006; P. D Magnus 2009; Fallis 2008; Rodriguez 2007). From cultural studies, issues of multiculturalism in Wikipedia have been compared with off-line multicultural interaction (Pfeil, Zaphiris, and Ang 2006). Wikipedia’s system of policy making and dispute adjudication has also been analyzed extensively with respect to existing legal systems and jurisprudential models (Benjamin 2006; Holland 2008; Henrion 2007; Geiger 2007). In fact, one the first academic articles about Wikipedia was an analysis of legal authority in various free knowledge projects, including the young Wikipedia (Benkler 2002).
In addition, there has been an extensive amount of sociological research into Wikipedia, which has provided detailed descriptions of the project’s social structure. Such research, taken as a whole, articulates the composition of the Wikipedian social structure, describing in various scopes and methods the way in which hundreds of thousands of participants coordinate their behavior in the shared task of collaboratively building an encyclopedia. The amount of research into sociological phenomena in Wikipedia is significant, far exceeding that performed by any other disciplinary perspective in the humanities or social sciences.

One possible reason why such a large amount of sociological research has been performed into Wikipedia is that common sense – which is the primeval form of sociological theory – states that a group of hundreds of thousands of volunteers could not possibly self-organize in such a way that would enable them to collectively write the largest and most accessible general-purpose reference work in human history. For such a feat to even be possible, it is claimed, there must be some organizing force – technical, social, governmental, cultural, or all of the above – that directs volunteers to act in certain ways and not others. However, this contradicts popular representations of Wikipedia as an anarchistic state-of-nature that is organized more like a mob than a productive community of knowledge production. While volunteers may come in to Wikipedia with vastly different and often contradictory intentions, it is assumed that some social structure must exist to bend their will into conformity in order to explain the project’s relative stability.

Operating from this assumption, many researchers have entered Wikipedia
and attempted to document such a social structure. Most conclude that while Wikipedian
society may seem alien or even anarchistic at a first glance, deeper comparison shows
that it is similar to various off-line organizations. Specifically, such research has
compared Wikipedia to groups that have successfully produced public or commonly-held
goods in a participatory, egalitarian, and consensus-based manner, including Quaker
communities (Reagle 2008) and participatory budgeting councils in Brazil (Navot 2008).
Another common trend has been based around Ostrom’s principles of collective action
(Ostrom 2000), which were developed through analyzing field studies of successful and
unsuccessful self-organizing groups organized around public or commonly-held goods.
Many researchers have claimed to have confirmed such processes at work in Wikipedia,
which are taken to explain why the project has not degenerated into a lawless anarchy.
Like the social order that maintains fish stocks of the Chisasibi Cree, community
irrigation systems in Spain, or forests in Japanese mountains, Wikipedia’s social structure
is said to work to maintain order in the free encyclopedia that anyone can edit because it
has developed and deployed the right combination of social norms, rules, roles,
institutions, and practices (Viégas et al. 2007; Viégas, Wattenberg, and McKeon 2007;
Butler, Joyce, and Pike 2008).

A common example given by sociologists is a common conflict in Wikipedia: an
edit war. Two (or more) editors get involved in a dispute over whether or not a certain
factoid ought to appear in an article: one removes the fact, which is put back by the other,
who removes it again, and so on. Yet because of the project’s social structure, this does
not turn into an endless anarchistic battle of the wikiwills, even when both
participants call on others to support them in keeping the article in their desired state. Instead, the participants and their allies are forced to debate the issue in a designated space according to various agreed-upon conventions, norms, and processes. When one position emerges as victorious in this conflict, that is, when a consensus is reached regarding the factoid’s place in the article, it is accepted as legitimate by all participants – even those who disagree with the outcome. Like in any society, it is claimed, this pre-existing normative infrastructure is what keeps the body politic unified, even in dissent.

In the sociological view, when individuals enter Wikipedia, they therefore enter a realm ruled according to a largely invisible set of social forces. Like lawyers or clerks in courtrooms, doctor and nurses in hospitals, and workers and managers in factories, editors must act in accordance with these social forces or else risk failure and even expulsion. While veterans are skilled in navigating and sometimes even further developing this system, the only way newcomers learn about these forces is when they brush up against them by performing some socially unacceptable act. And, as polemical, anecdotal accounts of Wikipedia by journalists and academics almost always claim, the criteria for a social acceptable act in Wikipedia is often radically different than those in other spaces of knowledge production.

**The Social Construction of Wikiality**

For critics of Wikipedia, the fact that the production of knowledge is dictated by what appear to be arbitrarily chosen social structures runs contrary to traditional understandings of knowledge production, especially when compared to the natural sciences. A perfect example of this is Computer Science professor Simson
Garfinkel’s claim that:

> With little notice from the outside world, the community-written encyclopedia Wikipedia has redefined the commonly accepted use of the word "truth." […] Unlike the laws of mathematics or science, wikitruth isn't based on principles such as consistency or observability. It's not even based on common sense or firsthand experience. Wikipedia has evolved a radically different set of epistemological standards … [in which] objective truth isn't all that important, actually. (Garfinkel 2008, 84)

As Garfinkel claims, the problem with this mode of knowledge production is that the social and epistemological model agreed-upon by Wikipedians brackets out truth as a criterion for knowledge and substituted what the project has called "verifiability, not truth." He attacks this standard as being "an appeal to authority--not the authority of truth, but the authority of other publications" and claims that "many publications don't do any fact checking at all" (2008, 85). It is not that Wikipedia fails to maintain order and is subject to petty vandalism, slander, or ideological infiltration; in fact, Garfinkel claims that the project’s epistemological model, codified into three core policies, works rather well to orderly structure the way in which knowledge is produced:

> These policies have become the social contract for Wikipedia's army of apparently insomniac volunteers. Thanks to them, incorrect information generally disappears quite quickly. […] These policies have made Wikipedia a kind of academic agora where people on both sides of politically charged subjects can rationally discuss their positions, find common ground, and unemotionally document their differences. Wikipedia is successful because these policies have worked (2008, 85).

Rather, the problem is how Wikipedians define a ‘successful’ resolution of an article dispute, which, as Garfinkel laments, has absolutely no relationship to the truth-content of its articles. While such a Habermasian paradise in which a group of equals freely express their views and work together to reach a consensus on the issues
may be an ideal way to resolve political disputes, the sociality of the situation is precisely what makes it inadequate for the resolution of hard, epistemological disputes. For Garfinkel, the consensus-based, discussion-tempered standard of verifiability, not truth leads to a system in which “truth is received truth: the consensus view of a subject” (2008, 86) as determined by “standards that aren't especially surprising given that the site is rooted in a Web-based community, but that should concern those of us who are interested in traditional notions of truth and accuracy” (2008, 84)

This criticism of Wikipedia is relatively recent, which may seem surprising given that so-called scientific realists have been denouncing Wikipedia as relativist since it was founded. However, there is a fine but clear distinction between Garfinkel’s critique and those initially launched against Wikipedia – which largely focused on its relativist instability. Before it was widely reported in late 2007 and early 2008 that Wikipedia rejected an 180,000 signature petition to remove a depiction of the Islamic prophet Muhammad from its articles, criticism of Wikipedia was largely based in the belief that Wikipedia was premised on a purely populist commitment to subjecting knowledge to nothing but the wavering whims of the masses. Even if Wikipedia was able to achieve and sustain its majoritarian ideals in the face of vandals and ideologues, so the criticism went, statements are presented to the public as true if and only if the masses agreed they are true, not because they are actually true.

When the Wikipedian community rejected the petition on the basis that it went against the project’s editorial policies, many began to reconsider their majoritarian view of Wikipedia and investigated these ‘policies’ and other mechanisms which
supposedly governed the production of knowledge in Wikipedia. Like the old critique, the new critique accuses the project of relativism, and it is also premised on the view that even if Wikipedia was able to achieve and sustain its ideals, its articles would still be faulty. However, it is more nuanced, taking into account the social processes which ideally anchor claims to truth: the argument is that in Wikipedia, statements are presented to the public as true if and only if the masses agree that they pass certain standards, and the only reason why these standards are in force instead of other, more ‘traditional’ ones is because they happened to have worked for everyone involved, and not because these standards verify that the statements are actually true. In fact, as Garfinkel claims, Wikipedia’s epistemological standards are nothing more than a formalized appeal to authority, which philosophers have long classified as a logical fallacy.

Garfinkel gives several instances in which Wikipedia’s formalized processes required that false statement not be replaced with a true one because there was a reliable source verifying the false statement and no reliable source verifying the true one. One well-publicized example is that of Jason Lanier, whose Wikipedia article stated that he was a film director – among other professions. This statement was verified in Wikipedia by a reliable source that, according to Lanier, was incorrect. When Lanier tried to remove this statement, he was consistently reversed by various editors, even when explained who he was and why the source was wrong. As Garfinkel explains, Lanier was not the most reliable source on himself according to Wikipedia’s standards, even though he ought to be considered the world’s foremost expert on the subject. Only after Lanier wrote about the incident in a newspaper – and included the fact that he was no longer a film
This genre of Wikipedia criticism, in which an expert of some sort bemoans the fact that their edits are consistently reversed, has been a staple of journalistic depictions of Wikipedia since the project entered the popular imagination around 2003 to 2004. The way in which experts express their frustration follows a similar narrative form: First, the expert tells of discovering some incorrect fact on a Wikipedia article, usually about him or herself. Next, the expert chronicles his or her attempts to remove the offending fact, which are almost always thwarted by established Wikipedian editors who claim that the expert has not followed policy. The expert almost always attempts to use his or her credentials as well as personal experience to make the argument, which is not taken to be a good reason by Wikipedians. Frustrated and often insulted, the expert gives up and writes an article about the experience in which Wikipedia is critiqued.

These accounts of experts who get involved in Wikipedia are strikingly similar to those of scientists who participate in “transcientific” debates, that is, debates which involve science but take place in a non-scientific context, such as politics, law, journalism, religion, etc. In the move from:

the relatively stable traditional forms of scientific controversy [to] other—equally stable—forms of disputation … scientists who are only accustomed to the scientific mode of disputation are not well prepared for the debating rituals of transcientific controversies. They bring into the proceedings the scientific expertise and presentational skills which have stood them well professionally and find that these do not work as usual (Ziman 2000, 197).

However, we must pay attention to the specific criticism of Wikipedia that is prevalent among experts in recent years. There has been a marked shift away from the
claim that the project follows some form of anti-expert relativist collectivism that panders to the masses and leaves it only a step away from mob rule. In its place, new criticisms like Garfinkel’s are emerging that still level the charge of relativism, but targets social norms and institutions instead of ascribing a postmodern populism to the project. The question that the popular media is increasingly asking is not whether Wikipedia is accurate – for that requires double-blind studies of experts – but how the project’s cultural and social aspects materially affect the production of knowledge, often in local cases regarding a single fact or editor. In these accounts, some aspect of the social structures designed to keep the project running is supposedly revealed to be open to undue influence. In the same way that reporters aim to reveal corruption in Congress or manipulation of the markets, these reporters strive to show how Wikipedia’s weakness is not blatant, one-off incidents of vandalism – say, when a picture of George W. Bush is replaced with one of Hitler – or incidents when an organization silently edits an article without anyone knowing. Instead, they try to show how individuals and organizations are able to authoritatively advance their own, often false views precisely because they are able to twist Wikipedia’s policies and procedures designed to ensure truth and maintain civility. The core this critique is that decisions about knowledge are made on some non-epistemological basis, such as an editor’s clout or ability to deploy jargon.

In its 13 February 2008 edition, the feature article in SF Weekly was “Wikipedia Idiots: The Edit Wars of San Francisco” (Spicuzza 2008), premised on the various editorial disputes over San Francisco–related articles in Wikipedia. However,
Spicuzza’s article is less focused on the disputes themselves and more on Wikipedia’s “disciplinary system in place to deal with edit wars.” In particular, the article focuses on one pseudonymous editor, Griot, who reportedly twisted such rules and procedures to his advantage. “Despite the potential for being blocked,” she wrote, “anyone who has peered into the history of Wikipedia entries knows violations of Wikiquette are rampant. Griot has been accused of many such violations.” In fact, she wrote that according to other Wikipedians, “Griot seems to have a skill for baiting others to the point where they violated Wikipedia policies. He would then report them to administrators, which could result in them being blocked.” In all, Spicuzza’s article described in rich detail how the policies and procedures meant to protect ideals of Wikipedia – “openness, inclusion, and contributing knowledge as you see fit in a collaborative way” – were used as weapons by Griot against his opponents to bypass the collaborative discussion process.

How Wikipedia’s policies shape the behavior of its editors was not a new question in the media, but it quickly became one asked and answered by many journalists. In its 8 March 2008 edition, The Economist led its Technology Quarterly section with “The Battle for Wikipedia’s Soul” (The Economist 2008) its third article featuring Wikipedia. The sixteen-hundred word article focused almost exclusively on the project’s growing conflict between inclusionism and deletionism, “two conflicting visions … at the heart of a bitter struggle” regarding the project’s editorial scope. In stark contrast to the egalitarian majoritarianism that had dominated journalistic portrayals of Wikipedia, the article described the “Kafkaesque bureaucracy” in place to resolve the ideologically-charged debates over which articles meet the project’s standards for
inclusion. The victors in these battles over knowledge were not those who were able to bring forth the most allies through politicking and electioneering, but rather those who are able to best deploy “volleys of arcane internal acronyms and references to obscure policies and guidelines.” Wikipedian administrator and journalism Professor Andrew Lih was quoted lamenting the “danger” of such “wiki-lawyering,” characterized as an unfortunate consequence of “the proliferation of rules, and the fact that select Wikipedians have learnt how to handle them to win arguments.”

Comments made by Larry Sanger, who was Wikipedia’s co-founder and self-proclaimed “chief apostate” also follow this trend. In a 2005 memoir, Sanger gave his account of the “early history of Wikipedia and Nupedia” (Sanger 2005a) the latter being the former’s failed predecessor, established to leverage the efforts of volunteer scholars in writing a free encyclopedia. In his account, he claimed that Wikipedia’s success and failure was due to the community of editors. In the beginning, the project was comprised of ex-Nupedian academics who established a good foundation for a wiki-based encyclopedia project. As Wikipedia experienced unprecedented growth and popularity, he claimed that the newcomers took the essence of the project to be one of anarchy, dooming the project to a cultural milieu that did not respect experts. However, this purely majoritarian view shifted when Sanger “forked” Wikipedia and founded his own wiki-based encyclopedia project that would be edited by experts and those who respected them, Citizendium. In addition to picking the right pool of contributors, he introduced new policies which strongly departed from Wikipedia’s in order to better achieve this goal. First, contributors had to register with their real names and electronically sign
a statement of principles. Soon after, Wikipedia’s policies on citing sources, neutrality, deleting articles, and dispute resolution were denounced as unscholarly and replaced. Finally, by 2007 the rhetoric shifted entirely as Sanger argued that the site “must resemble a law-governed, constitutional republic” (Citizendium 2007) and deemed there were to be systems of rights, representation, and justice.

This same shift has even occurred in Andrew Schafly, best known as the founder of Conservapedia, a wiki encyclopedia that was founded to correct the perceived “liberal bias” of Wikipedia. When the site was founded, Schafly explained that Wikipedia’s bias resulted not from its principles and ideals, but rather the ideological makeup of its editors – who were on average six times as liberal as the average American according to a Conservapedia study. According to him, Conservapedia was founded to do “what Wikipedia says they are trying to do but actually don’t do” (Zeller 2007) by establishing a wiki encyclopedia with a conservative community of editors but otherwise identical to Wikipedia. This argument – that the liberal bias of Wikipedians led them to ignore their pledges to presenting verifiable information in a neutral point of view – was premised on a majoritarian view of Wikipedia common to journalistic accounts of Wikipedia before mid-to-late 2007.

However, Schafly’s criticism of liberal bias in Wikipedia shifted from the ideological makeup of its users to the project’s policies and ideals. For Conservapedians, it became clear that Wikipedia’s ideals were not corrupted by the liberal community, but were themselves hidden sources of liberalism designed – explicitly or otherwise – to foster an anti-intellectual climate that suppresses and marginalizes true
knowledge, especially that which contradicts liberal ideology. For example, in an essay on how “How Conservapedia Differs from Wikipedia” (Schlafly et al. 2007) it is claimed that Conservapedia lacks Wikipedia’s “exacting set of rules which are designed to suppress original thought, new ideas and penetrating insights.” Likewise, they claim that adhering to Wikipedia’s Neutral Point of View standard would require them to support terrorism; they are instead “neutral to the facts. If a group is a terrorist group, then we use the label ‘terrorist’ but Wikipedia will use the ‘neutral’ term ‘militant’.” Even Wikipedia’s style and formatting guide is seen as a norm which “implicitly encourages … long-winded, verbose entries, making it difficult to recognize the essential facts.” In place of these policies, Conservapedia is governed by rules which “do not allow liberal bias to deceive and distort”—such as a ban on excessive discussion of article content and other “unproductive activities” as well as a requirement that all article content be true, not merely verifiable as in Wikipedia.

These five criticisms of Wikipedia come from vastly different sources: Garfinkel, a computer science professor, published his article in MIT’s Technology Review; Spicuzza, a San Francisco freelance journalist, published hers in a paper distributed by Village Voice Media; Schafly, a homeschool teacher and self-described evangelical Christian conservative, published most of his criticisms on Conservapedia; Sanger, Wikipedia’s co-founder who holds a Ph.D in Philosophy, published his criticisms of Wikipedia on Citizendium as well as various technology blogs and websites; finally, while The Economist does not publish bylines, the newspaper’s editorial line has a reputation distinct from the other four sources. However, they are all indicative of
a shift away from a conception of Wikipedia as an unstructured, anarchic space of mob rule where knowledge is dictated by the collective whims of its contributors and towards one that critically interrogates the way in which the social order present in Wikipedian society materially affects the way in which knowledge is produced, contested, and sometimes even destroyed. In the language of the so-called ‘Science Wars’ of the 1990s, this critique is that knowledge in Wikipedia is a social construction that has no more than a tangential connection to truth, and should therefore be disregarded in favor of more traditionally organized institutions of knowledge production.

The problem with such a critique is the same one that social constructivists had when denying the validity and objectivity of ‘hard’ scientific facts. Like in science, epistemic outcomes Wikipedia may seem rather contingently decided when one examines its internal processes; however, Wikipedia seems to work relatively well in tracking the truth when compared to professionally edited encyclopedias. In fact, one of the most publicized categories of research into Wikipedia is evaluations of knowledge quality, usually compared to an authoritative reference work by blind experts. After the widely-discussed study by the editorial board of *Nature* (Giles 2005), which concluded that encyclopedia articles in Wikipedia and Encyclopedia Britannica contained roughly the same number of errors, a small number of similar studies erupted, spearheaded by both academics and journalists. While Wikipedia has not been shown to best all other encyclopedias in all cases or be 100% factually accurate, these studies (Stvilia, Twidale, Linda Smith, et al. 2005; Chesney 2006; Hu et al. 2007; Halavais and Lackaff 2008; Wedemeyer 2008) have generally shown that Wikipedia’s quality remains high.
What is most interesting about most of these studies is surprise that such expert reviewers often report when actually reading a Wikipedia article – while it may not be the best article or completely free of errors, Wikipedia’s quality remains far greater than what many expect to result given its editorial model.

Such findings, which are congruent with the previously-cited sociological research regarding Wikipedia’s social order, are difficult to reconcile with anecdotal and popular accounts of Wikipedia’s systemic bias and due to its social structure. The open question in popular and academic circles is whether Wikipedia’s social order holds together a community successfully aiming for truth or an easily-manipulated, systemically biased bureaucracy producing disinformation and bias. This issue is all the more complicated by the fact that equal and opposing claims of widespread ideological manipulation are leveled at Wikipedia. When the cultural critic – either in academic or journalistic form – asks whether or not ‘we’ should trust Wikipedia given the project’s social and cultural composition, the response is inevitably contradictory when all information about the encyclopedia project is taken into account. Wikipedia appears to be simultaneously anarchistic and orderly: it is a decentralized bureaucracy, an egalitarian oligarchy, a transparent cabal, a participatory technocracy, a groupthinking collective fractured by petty disputes, an epistemic free-for-all governed by editorial standards, and a cult of amateurs that proves the wisdom of crowds.
The Rise and Fall of The Wisdom of Crowds

At this point, it may seem that sociology has steered us wrong; that our sociological and common sense theories as to how people socially organize in modern life are at best in need of revision and at worst disproven by Wikipedia’s mysterious order. One group that adheres to this view are the followers of James Surowiecki, whose principles of collective decision-making in The Wisdom of Crowds (Surowiecki 2004) famously argued that social science needs to change the way in which it has traditionally understood crowd behavior. Followers of Surowiecki assume that Wikipedia’s surprising wisdom can be explained not because of any sociological factors present in its structure or cultural milieu, but because of economic and mathematical theories.

Surowiecki argues against traditional notions of organizational theory, holding that attempts to coordinate behavior in a group are inefficient and lead to worse decisions. Instead, he claims decisions should be made by the raw aggregation of individual opinion, the ideal vision of this being a stock market. His principles of wise crowds are largely based on the set of individuals involved in the decision-making or fact-finding process: as the size of a group grows, its members become increasingly more accurate in their collective choice, as long as group members are sufficiently diverse in their viewpoints, independent in their determinations, and able to draw information from local sources. While these qualities of group members form the core of Surowiecki’s theory of crowd wisdom, his fourth principle does state that there must exist some form of “aggregation” for decision-making crowds – wise or foolish – to even be
possible. He describes aggregation such that the concept is binary, that is, members’ decisions are either transformed into a collective decision or not: it only matters if “some mechanism exists for turning private judgments into a collective decision” (10). If a crowd’s members “satisfies those conditions,” Surowiecki claims, it is “a mathematical truism” (10) that they will make wise decisions or determinations, adding that “You could say it’s as if we’ve been programmed to be collectively smart” (11).

Surowiecki only discusses the concept of aggregation in a few short sections, one of which discusses “information cascades” (63), in which a would-be wise crowd is made foolish because of the aggregation mechanism: individual decisions are announced publicly and in sequence. In this situation, he claims, later members are influenced by the decisions of the early members instead of making their own independent choices; only when members have incentives to distrust others can a crowd be wise. Because of this, he claims that “One key to successful group decisions is getting people to pay much less attention to what everyone else is saying” (65). Later, in a related example, he argues against “consensus-based discussions” (221) such as a meeting of a board of directors, as the aggregation mechanism can be influenced by factors like rhetorical skill, negotiation or bartering, personality traits, and seating arrangements. In order to ensure that members make decisions solely based on the information available to them, he suggests that aggregation of opinions on issues “should be done informally and privately prior to any meeting” (222).

Another section advances the idea that aggregation is a form of centralization without “central planners” (75). He makes this point in discussing the failure of U.S.
intelligence agencies to predict the September 11th terrorist attacks. In his theory, aggregation – or rather a lack thereof – explains why the massive crowd of intelligence analysts failed to successfully make the right determination. Because of the various bureaucratic structures in place, the aggregation mechanism which culminated with a report to the president did not properly take into account the determinations of all members. In fact, he claims that such a system inhibited the independence of analysts, who were not able to freely make determinations because of organizational and political pressure. His treatment of aggregation in this case indicates that the concept is not binary, as there was a mechanism for arriving at collective decisions, albeit one that worked rather poorly at leveraging the wisdom of the crowd.

In discussing a new aggregation mechanism, Surowiecki rejects proposals for a new centralized organization to take over all aspects of U.S. intelligence, because “everything we know about cognition suggests that a small group of people, no matter how intelligent, simply will not be smarter than the larger group” (78). Instead, he suggests that aggregation could be achieved in part “simply linking the computer databases of the various agencies” or developing a “single unified ‘watch list’ that draws on data from all parts of the intelligence community” (78). However, in order to fully realize the collective wisdom of the intelligence community, he supports a prediction market in which agency analysts, upon reviewing classified information, could buy and sell wagers for various events. This is Surowiecki’s ideal form of crowd wisdom, and is an exemplar of his rational choice-based theory of crowd wisdom at work:

an internal market would help circumvent the internal politics and bureaucratic wrangling that have indisputably had a negative effect on American
intelligence gathering, in no small part by shaping the kinds of conclusions analysts feel comfortable reaching. In theory, at least, an internal market would have placed a premium not on keeping one's boss or one's agency happy (or on satisfying the White House) but rather on offering the most accurate forecast. And since it would have been open to people from different agencies, it might have offered the kind of collective judgment that the intelligence community has found difficult to make in the past decade. (79)

A Surowieckian crowd is primarily constituted around an ideal aggregation mechanism operating according to an individualistic, incentives-based, rational choice model. Despite the fact that he dedicates such a short amount of time in his book to the concept of aggregation proper – choosing instead to focus on the ideal qualities of crowd members – much of the book is implicitly focused on which aggregation mechanisms nurture these ideal member qualities. His theory is based on developing an environment in which many diverse individuals can enter a crowd with local information, weigh options based only on their interpretation of the information they have obtained, be free from not only coercion but any external influence whatsoever, be given incentives (usually economic) to both participate and make correct decisions, and have their individual choices be automatically formed into a collective decision.

Interestingly, Surowiecki does not cite Wikipedia at any point in his work; however, The Wisdom of Crowds both as a book and a phrase is routinely invoked to explain the encyclopedia project’s unexpected order and productivity. Yet despite this tendency to ascribe a spontaneously-emergent, economic-based order to Wikipedia, the project community exhibits none of the characteristics that Surowiecki describes (which is potentially why he does not mention it in his work). Local information that cannot be confirmed by reliable sources is rejected as original research. Members are hardly
diverse in any sociological measure, and they extensively discuss issues and attempt to arrive at a consensus. Information sharing is a key mode of discussion and argumentation instead of something that is structurally inhibited. Formal policies and informal expectations are frequently deployed to maintain order, and those who excessively violate policy or consensus are routinely banned. There are few economic incentives to participate at all, much less for making correct decisions. Finally, the formation of a collective decision is a complex and often lengthy process which must be constantly negotiated. However, this does not mean that the project is not a wise crowd or even a crowd at all. Nor is it the case because Wikipedia supposedly lacks aggregation mechanisms for transforming individual opinions into collective decisions. Overall, wise decisions do get made in Wikipedia, largely made on the basis of many individual choices instead of a single authoritarian declaration.

Existing research into Wikipedia has also called into question the applicability of his vision of a crowd in which independent individuals make determinations free of external influence, although some of his principles have been supported empirically. There have been many studies of Wikipedia that aim to test Surowiecki’s principles, and such research either focuses on editors or articles, leading to different research questions. Those focusing on articles ask if quality can be explained by the extent to which groups of editors contributing to particular articles fulfill Surowiecki’s requirements. Such studies have shown that the number of distinct editors to an article (Wilkinson and Huberman 2007) and the diversity of opinions in such a group (Arazy, Morgan, and Patterson 2006) are strongly and positively correlated with article quality. Others have
found that such factors only increase article quality when the members coordinate their actions with each other. In fact, article quality has been shown to decrease if participants act in a purely uncoordinated fashion, contributing independently without regard for the decisions of others (Kittur and Kraut 2008).

Studies focusing on editors ask if the group of individuals contributing to Wikipedia as a whole satisfy Surowiecki’s requirements for members in wise crowds. This kind of study examines the Wikipedian community as a whole and derives cultural or sociological measurements which are then mapped onto measurements of individual independence, locality, and diversity. These studies have shown that the size of Wikipedia’s editorial base has been increasing (Voss 2005), as has diversity and independence, evidenced by a “dramatic shift in the distribution of work to the common users, with a corresponding decline in the influence of the elite” (Kittur, Pendleton, et al. 2007, 8). Other studies of this sort call into question a Surowieckian explanation of Wikipedia, claiming that there is no automatic or technological aggregation mechanism to determine article content (Fallis 2008) and that editors are not diverse, as they are generally younger and more likely to be male, students, and English-speakers than the general population (Rosenzweig 2006; Nov 2007; Schroer and Hertel 2007). Finally, studies have shown that participants are expected to discuss and coordinate with each other, and that editors are regularly admonished for independently making edits without regard to community consensus (Stvilia, Twidale, Gasser, et al. 2005; Viégas et al. 2007; Kittur, Suh, et al. 2007). This incongruence with Surowiecki’s theory of crowd wisdom is due to the fact that in Wikipedia, order is not produced and maintained through a
stock-market like mechanism of opinion aggregation.

As frequently invoked as it is, Surowiecki’s theory fails to explain Wikipedia — and does not attempt to. However, another often-cited book does attempt to extend such an economic theory to the free and open encyclopedia project: Infotopia by Cass Sunstein (Sunstein 2006), who builds on Surowicki’s economic concept of crowd wisdom. Specifically, Sunstein is interested in the conditions under which crowds or networks can make accurate predictions without succumbing to a groupthink collective. For him, neoliberal economic theory may have some flaws, but it is far better than a structured, deliberative model of organizational knowledge production — which he typifies in the failure of the CIA to predict 9/11 and NASA to prevent the Challenger disaster. Instead, the market is a far better model for knowledge production, as he describes:

How might we obtain access to the knowledge that is held by many minds? One possibility is to build on the price system. As emphasized by socialism’s greatest critic, Friedrich Hayek, the price system is a “marvel,” simply because of its extraordinary power to aggregate information. As Hayek saw, markets produce prices for steel, books, coffee, and candy in a way that incorporates the dispersed information held by numerous people. In markets, people have an extremely strong incentive to get it right. Certain information may remain “hidden” in deliberating groups, but if there is profit to be had, consumers and investors will act on that information, which will not remain hidden for long. Partly for that reason, market prices typically reflect a massive amount of accurate information. They even create a kind of Daily Us. (14-15).

As such, nearly every explanation Sunstein gives for Wikipedia's success is dependent on the size and constitution of its human editorial base. As the wiki model simply aggregates the opinions of the users as they contribute to the encyclopedia in various ways, the project highly dependent on the kinds of people who edit it. Such descriptions are typical of many social network theorists, who predominantly focus on the
kinds of human agents at work:

large numbers of knowledgeable people are willing to participate in creating Wikipedia, and whatever errors they make usually receive rapid correction, simply because so many minds are involved. The involvement of many people ensures that Wikipedians are able to produce a much more comprehensive resource than a small group could, even a small group of experts. (151)

In particular, Sunstein’s explanation of Wikipedia’s resistance to vandalism shows the extent to which the project is entirely dependent on its editorial base. He argues that "Wikipedia works because the vandals are hopelessly outnumbered by those who want to make the project work." Another quote from Sunstein makes this position quite clear:

In this light, we can easily imagine a society in which Wikipedia would not work ... in which many or most contributors to Wikipedia are confused, error-prone, partisan, or eager to engage in vandalism. Here the wrongdoers would triumph, creating error and confusion or worse. The good-faith contributors would be overwhelmed. (159)

For Sunstein, as long as there are enough intelligent, well-informed, and level-headed humans actively editing Wikipedia in good faith, the encyclopedia project (like any crowd/network) will remain successful by excluding the minority of those members who happen to be poor editors for whatever reason. However, Wikipedia will fail if its editors were not so properly constituted – much in the same way that economists explain market failures as stemming from information asymmetries, collusion, corruption, or even Alan Greenspan’s famous indictment of “irrational exuberance” among economic agents.

Yochai Benkler’s *The Wealth of Network* (Benkler 2007) presents a more tempered approach to neoliberal economics and has a more nuanced theory of networked interaction. For Benkler, the emergent properties of a network is largely due to the
ability of any user to interact with any other user in the network: in the same way that an ideal market exists when any buyer can transact with any seller, an ideal form of knowledge production exists when any editors can reverse any other editor’s contributions. In economics, the multitude of these interactions at micro-levels leads to rational economic agents arriving at the best possible price of a good (i.e. one that will clear the market), unless certain externalities result in an inability for such transactions to freely occur. As such, a significant problem for neoliberal economics is transaction costs, the time and effort it takes to engage in a process of negotiation at such a level among all potential buyers or sellers. Because of the seemingly-seamless process through which nearly any user can connect to any other user and instantly make a transaction, what Benkler calls “networked economies” are far more flexible than traditional monetary economies based on exchange.

Benkler is critical of purely economic market mechanisms (which he claims are too constricting), although his networked economics is simply a version of neoliberal economic theory that does not require monetary exchange. Because of this, his study of networks has a near-exclusive focus on the individual autonomy of atomistic, uncoordinated human agents – as his frequent quoting of the Enlightenment-era philosopher John Locke would seem to imply. Other humans and non-humans are not relevant in this model unless they coercively restrict the ability of each particular agent to fulfill his or her desires, which in this framework means the ability to ‘transact’ with any other agent in the network. As such, Benkler is predisposed to ignore the processes through which individual action is made aggregate in relation to other agents –
particularly those which are non-human. While his analysis of networks in general by no means ignores the technical, his autonomy-based framework is entirely premised on the restriction of information flows to or from individuals. Specifically, he is involved in the way in which decision-making processes can be coerced through the constitution of technological systems:

Conditions that cause failures at any of these junctures, which place bottlenecks, failures of communication, or provide opportunities for manipulation by a gatekeeper in the information environment, create threats to the autonomy of individuals in that environment. The shape of the information environment, and the distribution of power within it to control information flows to and from individuals, are, as we have seen, the contingent product of a combination of technology, economic behavior, social patterns, and institutional structure or law. (147)

Because of this, Benkler is able to deliver a spectacular critique of “policy routers” – information systems that allow Internet Service Providers to restrict or shape the Internet traffic according to various algorithmized rules. The most devious case, Benkler notes, is when a large ISP in Canada made their customers unable to access the website of the Telecommunications Workers Union during their 2005 strike. Because of the near-monopolization of the ISP market, he argues that such private action is an unprecedented threat to consumer autonomy. However, when there are multiple choices for consumers (some filtered and some unfiltered), Benkler resolutely claims that the autonomy is inherently preserved. Ignoring the apparent fact that an unfiltered Internet connection may be far more costly than an unfiltered one, Benkler states that if consumers have both options and knowledge of such options, “the very choice to use the manipulated pipe can be seen as an autonomous act” (148).
As this example makes clear, Benkler’s single-agent framework does allow an analysis of technology, but can only focus on the extent to which they actively inhibit the autonomy of individuals by restricting otherwise free information flows. The shortcomings of Benkler’s neoliberal economic-inspired framework, with its focus on the rational choices by autonomous agents, become clear when he turns to Wikipedia later in his book. Like many other social network theorists, Benkler focuses significantly on the human actors which the software architecture merely connects, leading to explanations of Wikipedia’s social order that take for granted the stability and unity of society. As he describes:

While the system operators and server host—Wales—have the practical power to block users who are systematically disruptive, this power seems to be used rarely. The project relies instead on social norms to secure the dedication of project participants to objective writing. So, while not entirely anarchic, the project is nonetheless substantially more social, human, and intensively discourse and trust-based than the other major projects described here. (72)

While blocking is far more prevalent than Benkler suggests (an average of 800 users are blocked each day), this is a minor factual disagreement. It is the second part—and specifically his invocation of norms—that is more troubling. Benkler seems to imply that once the technological infrastructure is deployed to enable a certain kind of interaction, the ‘transactions’ that emerge are entirely social, contingent on associations between humans. As such, he posits a set of pre-existing and agreed-upon norms to explain why social order emerges in Wikipedia, which he admits seems counterintuitive given that anyone does have the ability to act in any way at any time. While the ties that bind editors are taken by him to be independent, temporary, decentralized, and essentially
social – which he praises because it gives the network flexibility – these same properties facilitate a micro-level process of encyclopedia-building around common ideals. For Benkler, it may initially be assumed that something like Wikipedia would only be possible if participation is limited to an already-existing community with a shared worldview; however, as he claims, Wikipedia does not need such a strong social order to facilitate collaboration:

What is perhaps surprising is that this success occurs not in a tightly knit community with many social relations to reinforce the sense of common purpose and the social norms embodying it, but in a large and geographically dispersed group of otherwise unrelated participants. It suggests that even in a group of this size, social norms coupled with a facility to allow any participant to edit out purposeful or mistaken deviations in contravention of the social norms, and a robust platform for largely unmediated conversation, keep the group on track. (73-4)

While Benkler is one of the few defenders to invoke norms in his explanation of Wikipedia, the underlying explanation of the project’s order is largely economic – albeit one that does not involve exchange, money, or property. For him, editing is a ‘largely unmediated conversation’ between a multitude of independent, uncoordinated individuals. He also pays attention to the software, but not beyond the ‘anyone can edit’ mechanism which lets any editor ‘transact’ with any other editor. As in an ideal free market, there are no intermediaries to make transaction costs skyrocket in Wikipedia, meaning that for Benkler, there only needs to be activity at the micro-level in order to produce coordinate effects (i.e. order). This is why incentives play such a large role in the work of Surowiecki, Sunstein, and Benkler, as they must exist in proper configurations to motivate freely autonomous individuals to act in ways that
maximize good behavior. In other words, the crowd environment must be set up such that a rational individual seeking only to maximize his or her own self-interests will be compelled to make decisions that are in line with the benefit of the group. In fact, the theory holds that virtues or instincts like coordination, altruism, and information sharing are actually harmful to the crowd’s decision-making ability and must be either made impossible or disincentivized such that members will be compelled to make individual decisions.

This assumption – that the only way in which a crowd can be made wise is a market-based mechanism with corresponding incentives – is latent in a significant amount of academic literature and popular accounts regarding Wikipedia, especially so with criticism of the project. One of the best examples of this is a comparison of epistemic cultures in science and Wikipedia (Wray 2009). According to Wray, science has been established such that the actions of individuals are guided by the actions of a libertarian “invisible hand” (40) which ensure that all members involved made decisions which lead them towards truth. As individual scientists seek recognition from peers and funding agencies, they must act in certain epistemologically sound ways in order to maximize their economic self-interest. In this view, it is only because the environment has been set up such that “personal interests coincide with institutional interests” that “science works so effectively” (42).

In contrast, Wray claims that the “ethos” (40) of Wikipedia is far different, such that there can be no possible invisible instrumental mechanism to guide the motivations of Wikipedian contributors towards truth. This is because Wikipedians “lack the sorts
of incentives that keep science in good working order” (43), such as peer recognition for
good contributions and alienation for bad contributions. In particular, he claims that
when Wikipedians make disastrous errors in judgment or flagrantly mislead the
community, “their livelihoods are not threatened, as is the livelihood of a negligent
scientist” (43). Instead, the only incentives that Wikipedians seem to have is the
“prestige attached to being the first to post facts” (44), which he claims makes Wikipedia
“a site for gossip” and leads to the project’s “puckish culture” (44). These cultural
differences, Wray claims, are obvious to “most of us … especially those of us who study
science” (45). However, he claims that the general public is largely not attuned to the
vast extent to which Wikipedia has “created a community of inquirers who are governed
by norms very different from those that govern other more familiar communities of
knowledge producers” (47).

Another much-publicized critique on similar grounds has been made by Eric
Goldman, who has continually predicted since 2005 that Wikipedia will fail by 2010. In
2005, he predicted that Wikipedia would be “taken over” by marketers and vandals,
killing the project’s credibility, decreasing incentives for its regular editors to resist such
incursions, and inevitably leading to “downward spiral” such that “it will lose all
credibility” in five years (Goldman 2005). While the project could choose to restrict
editing in order to prevent the collapse of credibility, this would cause a community
backlash and decrease member incentives to the same effect. In 2006, he renewed his
prediction of the “death spiral” caused by decreasing incentives for Wikipedians and
claimed that “we can see the very beginning of the attacks that will lead to”
such a total collapse (Goldman 2006). By 2008, he was still making the same claim, arguing that the project’s “xenophobia” is evidence that it is at a crossroads where it “must choose between being high-quality or freely editable” although both options, he argued, will lead to the supposed “death spiral” by removing incentives for genuine participation among its established editorial corps (Anderson 2009). In this view, the project will fail because its model of incentives for good editorial behavior is not sustainable. Other critics who have made similar arguments against the sustainability of Wikipedia have supported alternative wiki-based encyclopedia projects that allegedly provide more or better incentives (some economic, others psychological or culture) for good editorial behavior and harsh disincentives for violations. However, Wikipedia still remains a far more orderly, active, and productive project than rational choice theory predicts – even with modifications like Surowiecki’s, which provide an incentives-based framework for explaining decentralized, collective action in the context of instrumental reasoning.

What, then, explains Wikipedia’s order and stability? Even if Wikipedia’s frequently-discussed norms regarding civility, assuming good faith, reliable sourcing, and neutrality would work if strictly followed, there is no reason that they ever will be, especially by anonymous users. If individuals have no rational reason or other incentives for acting in socially and epistemologically amenable ways, common sense tells us that many of them will act in whatever way they desire. As anyone who has worked in an office with a common refrigerator knows, norms by themselves are poor enforcers of behavior: if there is little risk of getting caught and a large enough reward, some
people will violate any number of well-established rules of common courtesy and private property, no matter how much moralizing has been attempted through signs and departmental memos. However, a staggering amount of sociological research has shown that Wikipedians do act in an orderly fashion in accordance with established norms in socially amenable ways. If Surowiecki’s theory does not apply to Wikipedia, then what explains why Wikipedia’s editors act this way?

To this end, many surveys of Wikipedians have been undertaken to try to answer this question. As may be expected, highly-active users report finding strong enjoyment in both the editing process itself and membership in the community of editors. Because of this, it has been frequently hypothesized that this desire for community acceptance explains the order of Wikipedia: given that users seek to improve or at least maintain their reputation and desire praise from fellow members, they will generally act in socially amenable ways to maximize their enjoyment. However, highly-active users only comprise a minor fraction of the project’s editorial base. In addition, the previously-cited study by Kittur et al. showed a dramatic decline in the amount of work performed by such “elites” at the expense of a growing group of infrequent contributors, and other studies have shown that anonymous users comprise a significant percentage of contributions to the site – between 25 to 30%.

While many of these anonymous edits are vandalism – 97% of all vandalism is by anonymous editors, according to one casual study by Wikipedians – there is significant resistance among Wikipedians to banning all anonymous contributions. This is largely because of the many anonymous “good Samaritans” (Anthony, Williamson, and
Sean Smith 2005) who contribute anonymously and infrequently yet collectively add much value to the encyclopedia. In two other causal studies by Wikipedians, “about 76% or 82% of anonymous edits are intended to improve the encyclopedia” and statistics generated by the Wikimedia Foundation have been cited as evidence that “A large portion of our good edits come from IP addresses” (Wikipedia Contributors 2009). This phenomenon is a mystery to both Wikipedians and sociologists, as Anthony et al. describes: “Virtually all theories of social dilemmas would predict low quality contributions from anonymous contributors, especially those with low levels of participation, since they would seem to have little motivation or incentives to contribute” (Anthony, Williamson, and Sean Smith 2005, 1).

However, the dilemma of anonymous and infrequent contributors extends far beyond their motivations for making edits. It is most likely the case that many individuals edit articles because they disagree with how a certain article is written, or wish to add their viewpoint to the article. As Wikipedia is one of the most popular websites worldwide, many non-Wikipedians likely have economic or ideological incentives to alter the content of various articles in given ways. In fact, many corporations and politicians have been caught anonymously editing Wikipedia articles in order to present themselves more positively or their competitors more negatively. Because of this, it should not be surprising that many anonymous users contribute in some fashion. The social dilemma is instead the fact that Goldman’s criticism has not come true – that anonymous marketers, ideologues, and vandals have not overrun the project with their bias. While the orderly, coordinated actions of highly-active
Wikipedians can be explained by the desire for community recognition and social praise, infrequent and anonymous users cannot be taken to have such motivations, especially when they may have conflicts of interests and incentives to forego coordination and edit against established norms and expectations.

Because of this, it must be assumed that Wikipedia is a dystopic society in the incentives-based rational choice view. As incentives are high for misbehavior among certain classes of editors (marketers, ideologues, vandals), it should be expected that disruptions of order are frequent, consistent, and inevitable. If anyone has the technical ability to edit any article in any manner, and a significant proportion of individuals have incentives to act maliciously or contradictory, rational choice theory predicts that chaos would ensue. The only possible explanation for Wikipedia’s order is that a staggering amount of social work must be done by the project’s ‘elites’ acting as vigilant police who seek out disorder and correcting it according to established norms, processes, and expectations. It is easy to see why Wikipedia is criticized by many economists, lawyers, political scientists, and analytic philosophers: such a society would be horribly inefficient, premised on ex post facto punishment and rectification instead of a laissez-faire environment in which individuals would freely choose to act in socially amenable ways without coercion. In order for Wikipedian society to be held together at a level higher than primal anarchism, these watchers must be constantly at work enforcing standards, norms, and expectations – both on ordinary editors and each other. Such critics often cite estimates that 30% to 40% of work in Wikipedia is dedicated to ancillary tasks like maintenance, negotiation, and governance, a figure that seems staggering
until compared with total government spending as a percentage of GDP for developed nations.

**The Normative Structure of Post-Production Control**

This view of Wikipedia – in which disruptions are corrected instead of prevented – has been characterized as “post-production editorial control” (Emigh and Herring 2005, 2) and is seen as the exact opposite of the mechanism by which a supposedly-wise crowd makes decisions. Wikipedia’s mode of aggregation is taken to be highly negotiated, social, and discursive instead of automatic, technological, and algorithmic – more like internal bureaucratic politics than Surowicki’s betting market for terrorist attacks. In fact, it is assumed that this socially-burdened decision-making system is a direct and often intentional consequence of the openness of the wiki software. It has also been theorized that this system relies extensively on a massive amount of labor by a small group of close-minded individuals. As such, researchers like Emigh and Herring claim that this undermines the democratic goals of the project:

> norms are enforced through the agency of dedicated, socially-approved members of the Wikipedia community [...] a few active users, when acting in concert with established norms within an open editing system, can achieve ultimate control over the content produced within the system, literally erasing diversity, controversy, and inconsistency, and homo-genizing contributors' voices. This is an unintended, and to our knowledge previously unnoted, side effect of the "democratic" affordances of wikis (Emigh and Herring 2005, 9-10).

In this conception of Wikipedia, coordination is achieved because such highly-active editors spend a significant amount of time patrolling articles and discussion spaces and correcting misbehavior according to norms, processes, and expectations that
they have constructed. However, the reason that Wikipedian society has advanced beyond primal anarchism is the construction of various social artifacts through which societal norms and expectations are codified and made durable. An exemplar of this perspective is presented by Stvilia et al. in their discussion on the coordination of information quality (IQ) work in Wikipedia. Their study examines the project’s largely social “infrastructure of quality evaluation and promotion mechanisms” (34), which includes codified standards, dispute resolution processes, social roles, and the software that enables communication. Specifically, they are interested in the role of this infrastructure in coordinating information quality work “despite what seems at first glance the seemingly anarchic operation of the project” (35). To this end, they advance a hypothesis that perfectly articulates what has become the dominant sociological view of social order in Wikipedia, although it is narrowly tailored to determinations of information quality instead of social order as a whole:

the presence of heavy contributors or power editors is essential to hold the community together and foster the application of common IQ standards and conventions. The majority of Wikipedia editors probably have never read the Wikipedia IQ criteria or style guides. Power editors, on the other hand — who in many cases, but not always, are administrators — are familiar with the IQ policies and norms. One might suggest that those power editors are the ones who follow the standards in their edits and promote and enforce their use in the community (Stvilia et al. 2008, 15).

In this view, social resources are developed and used by highly-active users to produce and maintain order in the community. For example, the authors describe how a set of criteria for featured articles was developed by a small group of highly-active users, which were then enforced upon other users in order to promote a certain expectation of
what a good article ought to be.

It is claimed that social artifacts like rules and roles have been created to make this work easier for such highly-active users, who can deploy them with force among most infrequent or anonymous users. Further studies have identified many other social resources through which participants are ordered into compliance by power users, a process that usually begins with discussions in various designated spaces. Most notably, many policies have been created, such as Neutral Point of View (NPOV) or Reliable Sources (RS), which contain agreed-upon conventions. When a user makes an edit that is not in line with these expectations – such as replacing a statement sourced by a scholarly article with a contradictory one sourced by a blog – an editor familiar with policies can rely on them when reverting the edit. In this way, the debate which may follow over such a reversion is framed around whether or not the edit complied with the Reliable Sources policy, and does not explode into debates about the philosophical possibility of true knowledge, alleged bias in academia and the media, the distinction between fact and opinion, the political question as to who ought to make such determinations, or any number of debates which can potentially occur when weighing evidence. In this way, such research assumed that coordination costs are significantly lowered and order achieved through such highly-active users constantly deploying pre-existing social artifacts (Stvilia, Twidale, Gasser, et al. 2005; Butler, Joyce, and Pike 2008; Beschastnikh, Kriplean, and McDonald 2008; Geiger 2008).

While the existence of such social artifacts may explain coordination between users who have similar motivations or desires for community approval, it does not
explain how order is achieved in the face of malevolent or disruptive users who have more incentive to advance their point of view than to coordinate their actions with others. As Butler et al. have identified in their study of Wikipedian policy use, this is a distinction between coordination and control: coordination “assumes that all participants in a community have the same motives and merely need to understand how to get something done” while control is based on “rules of prevention of behaviors that will disrupt the process of the organization” (Butler, Joyce, and Pike 2008, 1107). In the case of blocking, for example, the authors quote from the policy to argue for such a pre-emptive role: because it is not “meant for ‘retaliation,’ but instead for ‘encouraging’ appropriate behavior,” it is claimed that the blocking policy “guides normative behavior rather than punishes deviance” (1107)

However, this contradicts the earlier realization that in the case of malevolent or disruptive users, Wikipedia does not operate according to a pre-emptive or deterrence model of social order which seeks to prevent such vandals from contributing at all. As was shown, the project is instead focused on seeking out and fixing disruptions after they occur, acknowledging that disruptive users are inevitable, frequent, and continuous. Further, the authors’ mischaracterize the policy’s stated intent, which is admittedly subtle. While blocking is intended to be used for prevention instead of punishment for its own sake, this does not mean that it is intended primarily to deter potentially malevolent users from editing inappropriately in the same way that advocates of the death penalty argue that it deters potential criminals from committing capital crimes. Instead of altering the means-ends calculation of potential editors by increasing disincentives
for bad behavior, blocking is intended to physically prevent disruptive users from editing for a period of time. The blocking policy is not primarily intended to make users act more sociably when they otherwise would not follow established norms; instead, it is to prevent users who have shown a blatant disregard for established norms from acting at all.

More problematic is the fact that the authors ascribe to policy a kind of agency that it does not have on its own:

Since control mechanisms ensure consistency between the goals and actions of the individual and those of the community, and since the goals of individuals are sometimes destructive, such as vandalism, a policy like Blocking prevents chronic disrupters from damaging the group or its efforts. Blocking is the term for the prevention of editing rights for those participants who refuse to support the goals of the organization. (1107)

As previously stated, norms alone – even when codified into rules, memos, and signs – do not go very far in compelling behavior on their own, especially among individuals who act instrumentally, that is, who act on the basis of achieving set desires in the most efficient manner possible. The blocking policy could only be ascribed the agency that the authors give it if it were the entity which actually prevented disruptive users from contributing. However, it does not perform this action; as they describe, administrators acting in accordance with the blocking policy are the ones who block such users. Yet this would not be the case even if the blocking policy had the kind of deterrent effect on users they claim, such that the mere possibility of a block would be enough to alter user behavior. If this were true, the policy alone would not be the only actor in this story, as a document which states that punishment will result for certain acts is
nothing without a infrastructure through which that threat is made credible. What would
deter the potential disruptor would be the entire normative and technological system
through which disruptors are routinely identified and punished, or, for Foucauldians, the
mere perception that such a system is in place. Either way, the policy alone does not act
to deter or remove anti-normative behavior, especially for disruptive users.

In response, further sociological research into Wikipedia has investigated policies
with this understanding, examining the normative infrastructure through which policies
are used to produce order. The outcome of this research is a significant body of
literature that articulates the social structure which makes possible normative
development and enforcement:

A policy is an otherwise ordinary page in Wikipedia which “encodes and explains
norms” (Beschastnikh, Kriplean, and McDonald 2008, 1) for encyclopedia
building. They come on the scene almost exclusively in discussions and debates
when they are cited by participants to justify their positions; administrators cite
policy more regularly than others. Additionally, their relative citation frequency
in discussions has stabilized project-wide among all classes of users in recent
years as they have become general resources for normative enforcement. This
indicates that among community members, there has been “strong socialization”
(2008, 7)

Such research has supported the view that policies are a key factor in preventing
Wikipedia from degenerating into a lawless anarchy, and policy citing in discussions is
the main mechanism through which these social norms are reinforced largely without the
use of force (Viégas et al. 2007). Citing policy provides a “micro-to-macro link between
social norms and the self-governance structure” (Beschastnikh, Kriplean, and McDonald
2008, 4) such that any user can effortlessly perform governance acts. However, policies
and their citations do not stand alone, as they only exist within a normative framework that determines the relative authority of each policy as well as the conditions under which a policy can be legitimately used in an argument. Although disputes are sometimes resolved by arguments not based in policy, such as threats, *ad hominem*, and references to previous conflicts, this “policy environment” (Kriplean et al. 2007, 1) is an essential stabilizing element of Wikipedia’s social structure through which consensus is built locally and a sense of community built globally. The communicative structure of consensus-based editorial dispute resolution as mediated through the Wikipedian policy environment roughly approximates a Habermasian ideal speech situation through which rational discourse is used to resolve disputes in a far more emancipatory and democratic manner, compared to competing forms of knowledge production (Hansen, Berente, and Lyytinen 2007).

This research has also found that governance work in Wikipedia is largely tasked with enforcing the project's policies that structure editorial and behavioral conflict, and that because of the nature of policy environment, governance is relatively decentralized and participatory when compared to other organizations of knowledge production, although bureaucracies and structures of authority do exist (Butler, Joyce, and Pike 2008; Beschastnikh, Kriplean, and McDonald 2008). The project's roughly 1500 administrators, elected at large by the community, have the technical and social authority to autonomously perform three main governance tasks without consultation: temporarily banning disruptive editors, deleting unencyclopedic articles comprised entirely of vandalism or nonsense, and locking controversial pages from editing to force
discussion and a cooling of heads (Stvilia et al. 2008).

In the project’s early years, this governance work was taken to be performed in a centralized fashion by co-founder and self-proclaimed “benevolent dictator” Jimmy Wales, who was the final author, interpreter, and enforcer of policy (Ciffolilli 2003). However, through the development of the policy environment and the administrative class, Wikipedia shifted to a more distributed form of governance around 2004 and 2005. This decentralization has continued in recent years, as evidenced by the increasing moral and social authority of administrators, who are stepping into new social and governmental roles – including many that are not directly based on their technical privileges (Forte and Bruckman 2008).

Finally, this literature base has investigated who performs this work and why it is performed at all. For casual editors, who contribute to the article writing process as “legitimate peripheral participants” (Bryant, Forte, and Bruckman 2005, 1) with little to no understanding of the Wikipedian community, project norms, and the various social roles that editors can hold, this administrative work is largely invisible. However, this is not the case with editors who have become full-fledged Wikipedians, as these editors hold a shared conception of the community, nuanced and often conflicting views of project norms, as well as their own social identity linked to such groups and roles (Bryant, Forte, and Bruckman 2005). Administrative maintenance work is not invisible for these veteran members, evidenced by the fact that those involved in Wikipedia’s bureaucracies regularly give each other symbolic awards for performing these tasks (Kriplean, Beschastnikh, and McDonald 2008). Further evidence of the valued
status of bureaucratic work in the Wikipedian social structure can be seen in the election of new administrators: participation in these governance and maintenance processes is more predictive of promotion to administrative status far more than any other factor, including direct participation in the article-writing process (Burke and Kraut 2008).

Therefore, we can see that, according to the existing body of sociologically-themed research of Wikipedia, order is achieved through a tripartite social system: the policy environment, which gives the norms form, content, and legitimacy; the editorial user base, who locally promote compliance to norms through policy citations in discussions; and the administrative class, who develop and codify norms as well as enforce their compliance through largely-invisible maintenance work like blocking editors and locking or deleting pages. There is also a fourth level which is not frequently discussed in detail, but often mentioned: the Arbitration Committee, an elected, court-like body that adjudicates disputes over the proper interpretation and enforcement of such norms, usually deployed when administrative interpretations of policy conflict.

In all, this social system is taken to explain the order of Wikipedia, both in cases of disruptive, malevolent users who refuse to follow norms as well as in disputes between Wikipedians who both act in good faith and in accordance with their interpretations of existing normative commitments. It also supports the view that this after-the-fact system of social maintenance must be sustained by a significant amount of labor performed by the editorial base and the administrative class, who must patrol articles and discussions for disruptions and spend time locally restoring order as well as sanctioning the offending editor. Various social artifacts of the policy environment, such as templates,
accounts of best practices, roles, and workflows, exist to make this process more efficient among the small, but dedicated group of Wikipedians who collectively determine how Wikipedia ought to be administered. Because of the pre-existing structure of the policy environment, coordination costs between users in the administrative class are significantly reduced such that they can easily create new policies or punish users who violate existing ones.

A key problem in these accounts that the role of technology is minimal, almost exclusively dedicated to explanations of how various tools and features reduce coordination costs. In most sociologically-themed research, technical tools and elements are only discussed to the extent that they allow users to perform actions and complete tasks they would have done normally, except at a much faster rate or more accurate ability. While these authors may admit that certain technologies “conspire to make the sort of peer-process we find in Wikipedia possible” (Viégas, Wattenberg, and McKeon 2007, 9), they tend to take technology as a neutral force-multiplier or cost reduction tool and focus much more on the effects that social artifacts have in producing and maintaining such a social order. Others simply state that these tools are in use and broadly define their intended function, the most detailed of which is by Viegas et al, which is reproduced here in almost in its entirety with respect to two technologies, watch lists and talk pages:

Wikipedia allows users to keep a “watch list” of pages they wish to monitor closely. When a page in someone’s watch list is modified, the user is notified via email. This is an effective means for topic experts and serious Wikipedians to scrutinize changes made to specific pages and fix acts of vandalism such as mass deletions. Watch lists function as alerting mechanisms for wiki communities. The Wikipedia community also sets up secondary pages that are devoted to the
discussion of issues surrounding the topics on “real” pages; these are sometimes called “talk pages.” They represent an attempt to separate what is “real” information from discussions about what should and should not be on the real page. (Viégas, Wattenberg, and Dave 2004, 576)

Other researchers have integrated discussions of technology with sociality, spurred by studies into the different kinds of editing work performed by novice users and seasoned veterans. Such research has shown that part of “becoming Wikipedian” involves gaining familiarity with such tools such that they become “deeply integrated into their routine activities on the site” (Bryant, Forte, and Bruckman 2005, 6). Such authors should be celebrated for detailing how these tools allow veteran users to perform a different kind of work compared to casual users. However, at the time of their study, these tools were limited to discussion pages, page histories, lists of recent or new changes, and watchlists – and interestingly, they discuss the page history feature elsewhere, but not in terms of member use. Still, their conclusions are incredibly useful in understanding the role of such technologies in transforming work practices and social dynamics:

As they moved from the periphery into full community participation, these Wikipedians have assumed responsibility for maintaining the integrity of some set of articles. The watchlist is a tool they use to carry out this important task. Lists of recent changes and new pages also serve to alert Wikipedians to what other community members have contributed. These elements in the interface all provide Wikipedians tools for surveillance of the community. Rather than being faced with hundreds of thousands of articles to sift through in their efforts to maintain a quality resource, Wikipedians can focus on reviewing their personal watch list, recent changes and new articles. (7)

However, a significant time has passed since Viegas et al. and Bryant et al. investigated these technologies of work practice in Wikipedia. As previously-cited,
later research has claimed that a dramatic shift occurred in the project’s social structure around 2004 and 2005, and preliminary research indicates that many of these tools are no longer in use, having been superseded by others. Despite this, subsequent studies of Wikipedia have not continued with this theme as to how technological tools transform work practices and social dynamics. Exceptions to this have been rare and usually have taken the form of ancillary or explanatory suppositions which are not explored in depth. For example, in the previously-cited analysis of the policy environment by Beschastnikh et al., the authors note that one policy in particular – WP:SIGN, which says that users ought to sign their comments in discussions – experienced a sharp increase in citations in January 2006, which they explain due to the fact that a tool was developed to automatically sign unsigned comments. While they admit that the issue “touches on an interesting design point: to what extent should enforcement of policy be encoded into the software” (Beschastnikh, Kriplean, and McDonald 2008, 6) they leave this normatively-framed question for future research.

Most articles discussing technological tools which explicitly perform some social function or are discussed as having some effect on the sociality of the project – generally written from the fields of Computer Science, Information Science, or Computer-Supported Collaborative Work – are proposing new tools (Viégas, Wattenberg, and Dave 2004; Buriol et al. 2006; Zeng et al. 2006; Cosley et al. 2007; Demartini 2007; Kittur, Suh, et al. 2007; Suh et al. 2007; Witte and Gitzinger 2007), few of which have been taken up by the Wikipedian community. As Kriplean et al. rightly points out, “if such tools are to help improve awareness … their authors need to pay attention to how
they might be designed for appropriation by the Wikipedian community” (Kriplean et al. 2007, 175). On the other side, research dedicated to the analysis of existing tools in Wikipedia is largely focused on vandalism-detection and does not discuss on the sociality of the technology. Instead, these works either propose new algorithms (Potthast, Stein, and Gerling 2008; Smets, Goethals, and Verdonk 2008; Adler and Alfaro 2007) or evaluate the algorithmic effectiveness of existing implementations (Priedhorsky et al. 2007). In all, existing research has largely focused on the ways in which technology in Wikipedia has done made the editorial process more efficient, transparent, and effective. There has not been much recent research asking how editorial work itself – and with it, social relations tightly integrated with the practice – has been transformed in the wake of such technological delegation.

There is one major exception to the foreclosure of technology in sociologically-themed accounts of Wikipedia: the public, “anyone can edit” nature of the wiki medium. In fact, nearly every work which discusses sociality in Wikipedia makes note of the fact that nearly anyone can edit nearly any article at any time, and many theorize about how this alters the kind of social interactions which occur and the dominant social structure in which the project’s users are enmeshed. The fluidity and transparence of wiki pages has been theorized to be a crucial factor in making the social world in Wikipedia radically different from other, more traditional knowledge production projects, changing, for example, notions of authorship (Ray and Graeff 2008) and discursive power relations (Pentzold and Seidenglanz 2006). Likewise, Emigh and Herring’s previously-discussed criticism of Wikipedia is premised on the claim that the radical inclusivity afforded
by the software system is a key reason why it has developed the *ex post facto* system of social governance and why such a system is allegedly dominated by a small group of highly-active contributors. Others claim that governance in the project is more democratic or participatory than others because social norms and processes are present on publicly-editable wiki pages, whose full history can be viewed (Braun and Schmidt 2007; Viégas, Wattenberg, and McKeon 2007; Viégas et al. 2007).

In all, this research tends to deploy the specific technical features of the wiki software against grand technological narratives like Surowicki’s. Such a view of Wikipedia holds that the software inhibits automatic, technocratic decision-making and forces participants to involve themselves in the kind of social and local consensus-based discussions that Surowicki despises. Each additional feature built on top of the “anyone can edit” functionality – like discussion pages, lists of recent changes, and revision histories – are not characterized as automating negotiation or removing discursive burdens. Instead, these software extensions are seen as making the process more complex: features which increase transparency or participation, for example, are taken to increase the number of users who can object, seen in a positive light because they inhibit fast-moving, group-thinking oligarchies. Yet one problem with this research is that it largely remains speculative and at a theoretical level, lacking an empirical grounding as to how these technologies actually increase transparency or participation.

The main problem with this research is that each of these technical features have been present in the project’s software since the first few months of its existence. This makes it rather difficult to trace out the way in which these artifacts have
specifically transformed editorial work and social dynamics. Such research frequently leads to broad, technologically determinist explanations of social order and can only remain at the hypothetical level. Research cannot specifically analyze the social effects of technologies which have been present in Wikipedia’s software from its inception without, say, setting up an experiment in which controlled subjects participate in wikis when various technological features are enabled or disabled. While something like this has been performed through the analysis of internal wikis used by businesses (Tapscott and Williams 2006; Mader 2007), the resulting conclusions largely apply to wikis used by tens, hundreds, or possibly thousands of users. In contrast, in a given month Wikipedia has around 700,000 contributors: 150,000 who are registered and 550,000 who edit anonymously. Because of this, it should not be surprising that Wikipedia violates, often blatantly, a significant number of the qualities that such research has deemed to be essential to the success of any wiki.

The only tool developed in recent years to receive significant attention has been Virgil Griffith’s WikiScanner, which traces anonymous edits back to their physical source and allows users to see which organizations, corporations, or governmental agencies have been editing various articles. However, WikiScanner is generally discussed not as a tool used by Wikipedians, but rather one that has “brought the issue and idea of social transparency to the forefront” (Suh et al. 2008, 2) of the public imagination in discussions of whether ‘we’ ought to trust Wikipedia or not. This is largely because WikiScanner is not used by Wikipedians in their daily lives; over a year before the site launched, the community had already developed mechanisms for
identifying anonymous users by institutions. The tools which Wikipedians actually use in their average, everyday tasks of building and maintaining an encyclopedia have not generally not been studied by researchers, and especially not in terms of their sociality.
Chapter 2. Bots

Introduction

This chapter examines the roles of bots, which are automated computer programs that can directly interface with the wiki software, capable of analyzing articles and edits as well as make edits of its own. While technological tools have generally been passed over by researchers of Wikipedia in favor of more social artifacts like policies and standards, bots have been especially neglected. Research mentioning these computerized editors at all discusses them in one of several ways: first, as tools that researchers of Wikipedia can use for gathering sociological, behavioral, and informational data (Ortega and Barahona Gonzalez 2007; Medelyan et al. 2008; Burke and Kraut 2008); second, as information quality actors (usually vandalism reversers) whose edit identification algorithms are described and effects quantitatively measured (Cosley et al. 2007; Demartini 2007; Potthast, Stein, and Gerling 2008; Smets, Goethals, and Verdonk 2008); and third, as irrelevant entities which the software treats as users, meaning that they must be excluded from data sets in order to get at the true contributors (Voss 2005; Hu et al. 2007; Beschastnikh, Kriplean, and McDonald 2008; Almeida, Mozafari, and Cho 2007; Arazy, Morgan, and Patterson 2006; Ortega and Barahona Gonzalez 2007). Researchers that have discussed the social significance of bots in and of themselves have only made tangential or speculative clams, as such work generally has not focused on the social roles of such computerized editors.

Most research in the third category rejects bots based either for no stated rationale
at all, or based on findings made in 2005 and 2006 that, at their highest levels, they only comprise about 3 to 6 percent of all edits to the site (Kittur, Pendleton, et al. 2007), or that they are largely involved in single-use tasks such as importing public domain material (Stvilia, Twidale, Linda Smith, et al. 2005). As such, they have been characterized as mere force-multipliers which do not change the kinds of work that editors perform. Stivia et al., for example, conclude their discussion of bots by describing them as one tool among other, more social artifacts (such as standards, templates, rules, and accounts of best practices) that are “continually created to promote consistency in the content, structure, and presentation of articles” (Stvilia et al. 2008). This position is because their discussion of information quality work, like most discussions of Wikipedia, is focused on the actions of human editors. In such a view, bots do not perform normative enforcement of standards. Rather, ‘power editors’ use bots – along with rules and templates – in the same way that a police officer uses a car, ticket book, legal code, and a radar gun to perform a more efficient and standardized form of normative enforcement. While the authors do reveal important aspects of Wikipedia’s infrastructures, they are largely focused on unraveling the complicated standards and practices through which editors coordinate and negotiate. Research into Wikipedia’s “policy environment” (Beschastnikh, Kriplean, and McDonald 2008) or various designated discussion spaces has operated on this same human-centered principle, demonstrating the complex and often “bureaucratic” (Butler, Joyce, and Pike 2008) procedures necessary for the well-functioning of the project.

What is most interesting with bots is that they have been made invisible not
just in scholarship, but in Wikipedia as well: when an user account is flagged as a bot, all edits made by that user disappear from lists of recent changes so that editors do not review them. Operators of bots have also expressed frustration when their bots become naturalized, that is, when users assume that the bot’s actions are features of the project’s software instead of work performed by their diligent computerized workers. In all, bots tend to generally be taken for granted, and when they are discussed, they are not largely differentiated from human editors. As in with any infrastructure, technological artifacts in Wikipedia have generally been passed over, even as they have been incorporated into average yet essential maintenance activities. While such a view may have been appropriate when it was first made – around 2004 and 2005 – significant developments in bot operation have resulted in a massive increase in the number and scope of bot edits. However, recent research into the project largely passes over bots, operating under the assumption that the role of such technological actors has not changed.

In all, existing research into bots either takes for granted bots’ work, or assumes that this work might have otherwise been performed by humans in an identical manner. This case study contests both views of bots by illustrating how the delegation of tasks to technological tools can transform the kind of work being done, as well as interrogating the role of such non-human actors in the negotiation between competing claims of how various aspects of the project are and ought to be. Following Kriplean et al.’s call to study the way in which technological tools are “appropriated” (2007) or taken up by communities, this case study examines what now seems to be one of the most uncontroversial bots present in Wikipedia: HagermanBot, which was superseded by
the more descriptively-named SineBot. This bot monitors every edit made to Wikipedia in discussion spaces, looking for comments which have not been signed. If it detects one such comment, it appends a signature stating, for example, “—Preceding unsigned comment added by Jimbo Wales (talk • contribs) 11:15, Jun 1, 2005.” This bot is one of the most active users to Wikipedia, having made almost 600,000 edits at a rate of over one-thousand per day; as a result, many users assume that their comments will be signed if they do not do it themselves – as its creator has lamented. However, the introduction of a bot that automatically and indiscriminately signs comments was controversial when it was launched in December 2006, even though this work had been regularly performed by humans for some time.

This chapter traces out the way in which the introduction of the comment-signing bot shattered the perception of a common worldview that Wikipedians held regarding the role of bots as well as the norms associated with ownership of comments and discussion spaces. While some thought that the bot was a welcome addition and praised its use, others, for various reasons, did not want bots to automatically sign comments. As the community came to negotiate the bots existence, they did more than discuss what functions ought to be present in the bot’s source code – although this was the cornerstone of the controversy. In order for this negotiation to occur, participants had to make their world common again by re-establishing a shared conception of how their project ought to be in the context of this new kind of moral enforcer. As Wikipedians on all sides worked to determine the bot’s place in their world, a technical feature with a corresponding social justification was advanced in such a way that settled the controversy. This
compromise became codified – both into software programs and social expectations – such that it allowed all sides to not merely accept the bot’s existence but also the worldview upon which it was premised. Through a significant amount of work by humans and non-humans, the social and technical artifacts became fused and packed into a black box that became a taken for granted aspect of bots by many in Wikipedia’s bot development community.

Next, this account is complicated by contrasting the case of a bot developer who, years later, refused to accept this standard and remake his bot accordingly. Objectors to the developer were unable to gain support in the bot development community and had to extend their arguments to administrative networks in Wikipedia, which had largely delegated the task of bot regulation to the bot development community. The argument that the developer was violating a long-established de-facto standard held little weight among administrators, largely because this expectation had not been formalized into an official policy. Those who contested the bot’s irreverence of the standard in administrative spaces were not able to successfully translate their frustration into an actionable violation of a documented norm. Only when an editorial right of non-interference from unnecessary bot actions was authoritatively articulated could this technical standard circulate in administrative networks.

**The Case of HagermanBot**

**A Task and a Request**

At 06:15 on 17 October 2006, user ZeroOne made his 4,072nd contribution to Wikipedia, editing the discussion page for the article on “Sonic weaponry.” Instead of
adding a comment of his own about the article, he merely appended the text
{{unsigned|71.114.163.227|17 October 2006}} to the end of a comment
made by another user about twenty-five minutes earlier [05:50]. When ZeroOne clicked
the Submit button, the wiki software parsed his contribution and transformed it into a pre-
formatted message. Together, the edits of 71.114.163.227 and ZeroOne added the
following text to the article’s designated discussion page:

```
{{unsigned|71.114.163.227|17 October 2006}}

Two minutes later [06:17], ZeroOne performed the same task for an unsigned
comment made by a registered user on the talk page for the “Pseudocode” article –
adding {{unsigned|Blueyoshi321|17 October 2006}}. About two hours later
[08:40], he spent twenty minutes leaving {{unsigned}} messages to the end of eight
comments, each made on a different discussion page. While ZeroOne could have
manually added the text to issue the message, this process was made standard and swift
because of templates, a software feature that enables users to issue pre-formed messages
using shorthand codes. One of the most famous templates is {{fact}}, which appears
to viewers as [citation needed], although one of the most used templates is {{cite}}, which
allows users to issue standardized citations. When editing articles, templates
are invoked using double braces and like many, the \{{unsigned} \} template can be customized with fields, which are separated with pipe bars. The code for templates are located at pages in the wiki prefaced by “Template:” and followed by their name: the unsigned template is located at Template:Unsigned, while the fact template is at Template:Fact. Among Wikipedians, they are generally referred to either using double braces or their page name.

While the existence of templates made ZeroOne’s work somewhat automated, he felt that it could be made even more automated. Two hours after [11:03], he edited the talk page for the page “Wikipedia:Bot policy,” a space for discussion about the rules in place for approving bots – computer programs that automatically or semi-automatically edit articles and discussion pages according to pre-defined algorithms. ZeroOne made his own comment to this page, writing:

---

**Suggestion: Auto-signature bot**

I have noticed that perhaps six out of seven anonymous users who leave comments on talk pages do not sign their posts properly. I have usually added the \{{unsigned} \} message after those posts when I have encountered them. However, this could be a job for a bot: scan the Recent changes list limited to the Talk space, and if a comment is made by an IP-address, check it for a signature and add one if necessary. Of course logged in users also forget the signature sometimes, and those could be checked too, if it doesn't take too much resources. Alternatively only check those users that have not created an user page yet, they are often new to Wikipedia and do not know about signing their posts. Is anyone with the skill/equipment up to this? --ZeroOne (talk | @) 11:03, 17 October 2006 (UTC)

---

Within the hour, another user replied to his request:
Soon after, ZeroOne re-posted his suggestion in this discussion space, dedicated to discussions about requests for new bots. Over the next few weeks, a few users mused about its technical feasibility and potential effects without making any concrete decisions on the matter. The discussion stagnated after about a dozen comments and was automatically moved into an archive by a bot named Werdnabot on 16 November 2006, after having been on the discussion page for fourteen days without a new comment.

In the next month, another user named Hagerman was hard at work in realizing ZeroOne’s vision of a bot that would monitor talk pages for unsigned comments and append the \{{unsigned}} template message without the need for human intervention, although it is unclear if Hagerman knew of ZeroOne’s request. Like ZeroOne, Hagerman had used the template to sign many unsigned comments; however, many of these were his own comments instead of ones left by others. On 30 November 2006, having apparently finished programming the bot, Hagerman registered a new user account for HagermanBot and wrote up a proposal the next day. He submitted it to the members of the Bot Approval Group (BAG), a committee tasked with reviewing bot proposals and ensuring that bots are operated in accordance with Wikipedia’s policies. Within 36 hours, it was approved by an administrator named Tawker and at 00:06 on 3 December, it began automatically appending specialized \{{unsigned}} messages to every comment that it identified as lacking a signature. The first day, 790 comments were
autosigned, and HagermanBot made slightly over 5000 edits over the next five days. By the end of December 2006, HagermanBot had become one of the most prolific users to edit Wikipedia in that month, outpacing all other humans and almost all other bots.

Responding to Objections

There were a few problems with the bot’s identification algorithms that made it malfunction in certain areas, increasing the amount of work by Wikipedians instead of decreasing it. A few users left messages on Hagerman’s user talk page notifying him of such errors, which he promptly fixed. However, some users were annoyed with the bot’s normal functioning, complaining that it instantly signed their comments instead of giving them time to sign their own comments after the fact. For these editors, HagermanBot’s message was “embarrassing,” as one editor stated, making them appear as if they had blatantly violated the Signatures guideline. Others did not want bots editing messages other users left for them on their own user talk pages as a matter of principle, and an equally vocal group did not want the bots adding signatures to their own comments. Hagerman stonewalled the first group with a technical defense: that the wiki software was set up such that a delay would erase any edits made between the original edit and when his bot added the signature. He also quickly accommodated those who did not want bots editing their own user talk page. As shown in a designated discussion page for his bot proposal, it took Hagerman less than twenty minutes to go from proposal to implementation on this issue:
While he placated those who did not want the bot editing comments left for them, the issue raised by the other group of objecting editors was more complicated. These users were, for various reasons, firmly opposed to having the bot transform their own comments. One user in particular, Sensemaker, did not follow what was claimed to be the generally-accepted practice of using four tildes (~~~~) to automatically attach a linked signature and timestamp, instead manually adding “-Sensemaker” to comments. HagermanBot did not recognize this as a valid signature and would therefore add the {{unsigned}} template message to the end, which Sensemaker would usually remove. After this occurred about a dozen times in the first few days of HagermanBot’s existence, Sensemaker left a message for Hagerman on his user talk page, writing:

HagermanBot keeps adding my signature when I have not signed with the normal four tilde signs. I usually just sign by typing my username and I prefer it that way. However, this Bot keeps appearing and adding another signature. I find that annoying. How do I make it stop? -Sensemaker

Like with the previous request, Hagerman initially responded quickly, agreeing to exclude Sensemaker within ten minutes of his message and altering the bot’s code fifteen minutes later. However, he soon reversed his position on the matter after another editor said that granting Sensemaker’s request for exclusion would go against the purpose.
of the bot, emphasizing the importance of timestamps in discussion pages. Sensemaker’s manual signature did not make it easy for a user to see when each comment was made, which Fyslee argued was counterproductive to the role of discussion spaces:

[...] In my experience, users who deliberately don't use standard sigs are often those how are also engaged in behavior that attempts to contravene one of the things that makes Wikipedia work -- openness. (I am not accusing the user above of this behavior!) All our actions here should be open and above board, and capable of verifiability checks by other editors. The re-creation of the exact sequence of events in a discussion is essential. You've got a great bot here, and I wish you luck with it. -- Fyslee 14:23, 5 December 2006 (UTC)

Hagerman struck his earlier comments and recompiled the bot to sign automatically sign Sensemaker’s comments, calling Fyslee’s remarks “Very insightful!” As may be expected, Sensemaker expressed frustration at Hagerman’s reversal and Fyslee’s comment:

[...] I politely request that Hangerman turn his bot off for me. I was under the impression that he wanted to provide a genuine service to users who forget to sign occasionally (we can all be a bit absent-minded until we make a habit of it) -not force something unwanted upon those who expressly ask not to have it. Therefore the ability to opt out would be proper and respectful to the editors. I am sad to notice that he is hesitating in this matter but hope that he will not turn my request down. -Sensemaker —The preceding unsigned comment was added by Sensemaker (talk • contribs) 15:10, 5 December 2006 (UTC).

Sensemaker also raised this comment in the discussion page for Hagerman’s bot proposal, echoing his position that there should be some way to opt-out of its services and that Hagerman was appearing to not comply. Sensemaker urged the members of the Bot Approval Group to not allow Hagerman to “force something upon people who expressly
ask to be excluded.” However, some participants in the discussion – which was comprised of BAG members, administrators, and other Wikipedians – had a hard time distinguishing between the two kinds of opting out: the first allowing users to control bot edits to their user talk pages, and the other allowing users to control bot edits to their own comments.

Fyslee, on the other hand, knew exactly what Sensemaker wanted; framing it as a request “to deviate from standard practice,” he explicitly called upon the Signatures guideline and the project’s ideals of openness and consensus-based decision-making as an ally in his argument. At this point, Hagerman claimed that “the real question here” was not whether Sensemaker ought to be able to opt-out, but “whether a signature without a timestamp should count as a signature.” However, Sensemaker’s response and Hagerman’s rebuttal indicated that their dispute was over more than the ideal form of a signature or their proper role in discussion spaces. It instead became a dispute about the morality of bot operation, that is, whether or not it ought to be acceptable for someone like Sensemaker to make certain kinds of comments, and if not, whether it ought to be acceptable for someone like HagermanBot to alter those comments in order to bring them more in line with community standards:
For Hagerman and his allies, the issue was based on long-standing and well-known distinction between policies and guidelines. As an article on “Policies and guidelines” stated as of 7 December 2006:

A guideline is any page that is: (1) actionable and (2) authorized by consensus. Guidelines are not set in stone and should be treated with common sense and the occasional exception. [...] A policy is similar to a guideline, only more official and less likely to have exceptions.

Unlike the vast majority of existing bots, HagermanBot was created to universally enforce adherence to a guideline – one that that Hagerman and his allies thought was essential to the smooth operation of discussion spaces and therefore the activity of collaborative encyclopedia-building. He portrayed the two opt-out systems (!NOSIGN! [...] It was standard practice that people could sign their articles anyway they wanted to. Now you want to change the standard practice so that this is no longer possible. The person who argues for a change should have the burden of proof. If he wishes to impose his changes on people who really do not want them and expressly say so he or she must make a very strong case that imposing this change on an unwilling subject is warranted. I fail to see any significant benefit of the change and strongly dislike having them imposed on me for reasons of personal preferences. [...] Once again, I would not try to impose a new order on you or force you to sign your name in a certain way if you expressly said that you did not want to do so. Please extend the same basic courtesy to me. -Sensemaker —The preceding unsigned comment was added by Sensemaker (talk • contribs) 21:08, 5 December 2006 (UTC).

[...] I realize that signing posts is not a rule, but merely a guideline (and considered good etiquette). If you don't wish to sign posts normally, you are free to break the etiquette. One of the features that I've mentioned on the bot description is that you are able to disable the bot on a one-time basis by placing !NOSIGN! somewhere in your edit summary. You will need to do this each time to override the automatic signature. However, I implore you to consider signing correctly to make your comments easier to interpret to other editors. Best, Hagerman (talk) 00:10, 6 December 2006 (UTC)
and <!--Disable HagermanBot--> not as ways for users to decide for themselves if they ought to abide by the Signatures guideline, but rather to keep the bot from signing particular comments that, according to the guideline, do not need to be signed. From a design standpoint, HagermanBot assumed total editorial compliance with the signature guideline: the two opt-out features were to ensure more conformity, not less, by allowing users to tell the bot when a signature would be unwarranted according to the guideline. Users who were opposed to the Signatures guideline in general could use the tedious feature to make the bot not enforce the guideline when they made comments, but Hagerman begged Sensemaker to not attempt to opt-out.

Yet for Sensemaker and other editors, it was not clear “who gave you [Hagerman] the right to do this,” as one anonymous user who contested HagermanBot exclaimed. Hagerman responded to such rights-based arguments by linking to his bot proposal, which had been approved by the Bot Approval Group – clearly able to enroll this committee as an ally in defense of his bot. In fact, it seemed that Hagerman had a strong and properly-aligned network of actors in defense of his bot: a growing number of enthusiastic supporters, the BAG, the Signatures guideline, ideals of openness and transparency, visions of an ideal discursive space, the {{unsigned}} template, and a belief that signing unsigned comments was a routine act that had long been performed by humans. Yet for some reason unknown to them, users began objecting to this regular, uncontroversial practice when HagermanBot began performing it. Many users who had previously left their comments unsigned or signed with non-standard signatures began to make themselves visible, showing up at Hagerman’s user talk page and other spaces
to contest what they portrayed as an unfair imposition of what they believed ought to be optional guidelines.

Some objectors staunchly opposed to any user signing their comments, bot or human, and took issue with the injunction to sign one’s comments using the four tilde mechanism – Sensemaker was one of these editors, although others did not want to use a signature at all. Another group did not want to see a bot universally enforcing such a norm, independent of their stance on the necessity of signatures:

I don't really like this bot editing people's messages on other people's talk pages without either of their consent or even knowledge. I think it's a great concept, but it should be an opt-in thing (instead of opt-out), where people specify with a template on their userpage if they want it, like Werdnabot, it shouldn't just do it to everyone. Just my two cents. --Rory096 01:36, 11 December 2006 (UTC)

In responding to such objections, Hagerman and his allies shared their vision of how discussion spaces were and ought to be, which placed strong moral emphasis on the role of signatures and timestamps in maintaining discursive order and furthering the ideals of openness and verifiability. Yet while those like Fyslee were clearly able to articulate such a conception and enroll many allies in its defense, this proved inadequate to settle the controversy. This is because those opposed to HagermanBot began to articulate a different view of how Wikipedia was and ought to be – one which did not directly contest the claims made regarding the importance of signatures, discussion pages, and communicative conventions. Instead, those like Sensemaker advanced an opposing view of how users – and especially bot operators – ought to act to each other in Wikipedia, a
view that drew heavily on notions of mutual respect:

Concerning your emphasis on the advantages of the bot I am sure that it might be somewhat convenient for you or others to use this bot to sign everything I write. However, I have now specifically requested to not have it implemented against my will. I would not force something upon you that you expressly said you did not want for my convenience. Now I humbly request that the same basic courtesy be extended to me. -Sensemaker

Yet for Hagerman and his allies, these objectors were categorically taken to be irrational, malicious, or having what Rich Farmbrough called “botophobia.” While this seems to be a pejorative description that would strengthen Hagerman’s position, it restructured the controversy and allowed it to be settled in Sensemaker’s favor. In entering the debate, Farmbrough argued that while Hagerman and his allies were entirely correct in their interpretation of the Signatures guideline, Hagerman should still allow an opt-out system:

On the one hand, you can sign your edits (or not) how you like, on the other it is quite acceptable for another user to add either the userid, time or both to a talk edit which doesn't contain them. Nonetheless it might be worth allowing users to opt out of an automatic system - with an opt out list on a WP page (the technical details will be obvious to you)- after all everything is in history. This is part of the "bots are better behaved than people" mentality which is needed to avoid botophobia. Rich Farmbrough, 18:22 6 December 2006 (GMT).

The next day, Hagerman agreed, and the conflict was seemingly settled:
An Unexpected Ally

While the opt-out list may seem like a concession made by Hagerman, it proved to be one of his strongest allies in defending HagermanBot from detractors, who were arriving in numbers to his user talk page and other spaces – even after the Sensemaker/Hagerman dispute had been settled. Most users left value-neutral bug reports or positive expressions of gratitude, but a small but steadily-increasing number of editors continued to complain about the bot’s automatic signing of their comments. The arguments made against HagermanBot were diverse in their rationales and allies called forth, ranging from complaints based on annoyance to accusations that the bot violated long-established rights of editors in Wikipedia:

Ok, I've implemented an opt out procedure. Thanks again, Hagerman (talk) 03:35, 7 December 2006 (UTC)

Excellent, an opt out option was all I was asking for. A sensible but persistent appeal to a gentleman's sense of decency and reciprocity almost always succeeds. I am glad it did so in this case too. I have done a simple test and it seems to have worked. Thank you. -Sensemaker —Preceding unsigned comment added by Sensemaker (talk • contribs) 16:19, 7 December 2006

Glad it's working out for you. Best, Hagerman (talk) 23:43, 7 December 2006 (UTC)
As seen in Hagerman’s reply to the first objection, a human ally was useful in rebutting the objections made against his bot: the Bot Approval Group, which had reviewed and approved the bot according to established protocols. The Signatures

---

**Who gave you the right to do this?**

It is not mandatory that we sign, AFAIK. Instead of concocting this silly hack, why not get the official policy changed? I suppose you effectively did that by getting permission to run your bot on WP. How did you manage that anyway? (I won't bother with typing the fourtildas).

It isn't a policy, however, it is a guideline. You can view its approval at Wikipedia:Bots/Requests for approval/HagermanBot. Feel free to opt out if you don't want to use it. Best, Hagerman(talk) 02:29, 5 January 2007 (UTC)

---

**Point of Bot**

People have the right to not sign their posts, as an ip I didn't and no one ever got mad at me for it, so what's the point of this bot???? Teak the Kiwi 18:15, 14 January 2007 (UTC)

For people who don't know how or forget. People who don't want to would often find their posts "signed for them" using the {{unsigned}} template. HB just saves users the chore of doing it. No big deal. Rich Farmbrough, 00:01 15 January 2007 (GMT).

And as for the "right" not to sign posts, that is absolutely incorrect: Wikipedia:Sign your posts on talk pages. John Broughton | 02:54, 15 January 2007 (UTC)

Well... that's a guideline, it's not a policy. Teak the Kiwi does have every right to leave their posts unsigned, but HagermanBot has every right to leave the {{unsigned}} template. Best, Hagerman(talk) 02:44, 16 January 2007 (UTC)

---

As seen in Hagerman’s reply to the first objection, a human ally was useful in rebutting the objections made against his bot: the Bot Approval Group, which had reviewed and approved the bot according to established protocols. The Signatures
guideline – including the distinction between guidelines and policies – was also invoked to justify HagermanBot’s actions, as shown in both examples. It would seem that these actors, who are generally taken to draw their legitimacy from a broad, project-wide consensus, would be the most powerful allies that Hagerman could deploy in support of HagermanBot’s actions and its vision of how discussion spaces in Wikipedia ought to operate. However, a much stronger ally proved to be neither purely human or normative, but instead technological: the opt-out list through which angry editors could be made to lose interest in the debate altogether. It is this last actor that was most widely used by Hagerman and his human allies, who began to routinely use the opt-out list to respond to a wide array of objections made against the bot.

---

**Hinderance not a help.**

The only thing this bot has done for me is interfere when I have accidentally left of my signature. Rimmeraj 08:18, 3 February 2007 (UTC)

If you don't like the bot, you are always free to opt out. Best, Hagerman(talk) 17:44, 4 February 2007 (UTC)

---

**about this bot**

This bot operates a bit fast. I forget to sign, but deside rememver a second later, it already signed. It would be easier for wikipedia if it could operate slightly slower, like if it auto signed after a minute if you forgot to sign.--Sefringle 19:57, 1 April 2007 (UTC)

Opt-out ... it's the way to go. :) The Behnam 19:26, 2 April 2007 (UTC)
The strength of the opt-out list was its flexibility in rebutting the objections from two kinds of arguments: first, the largely unarticulated claims that the bot was annoying or troublesome to them; and second, the ideological or rights-based arguments that the bot was acting against fundamental principles of the project’s normative structure.
first argument was easy to rebut, given that the opt-out list completely responded to their objections, which were also relatively weak in the actor-network sense due to their inability to rely on allies in support. In contrast, those making the second kind of argument called forth a strong yet invisible non-human actor: juridico-political concepts of rights, autonomy, and freedom. Yet the same opt-out list could be invoked in HagermanBot’s defense against these actors, as it foreclosed their individual claims that the bot was violating their editorial rights. While objectors would have preferred that the bot use an opt-in list to ensure the rights of all editors, the opt-out list allowed HagermanBot to be characterized as a supremely respectful entity that was, as Hagerman’s new philosophy of bot building held, “better behaved than people.”

**Blackboxing the Compromise**

HagermanBot’s two new features – the opt-out list and the `<!--Disable HagermanBot-->` tag – soon became regular players in the Wikipedian community, especially among the bot development community. Rich Farmbrough saw the value of these non-human actors who helped settle the HagermanBot controversy, and wanted to extend such functionality to other bots; however, its idiosyncratic mechanisms were unwieldy. About a week after HagermanBot implemented the opt-out list, he was involved in a discussion about a proposed bot named PocKleanBot, which was described by its operator PockingtonDan as a “nag-bot” that would leave messages for users on their talk pages if articles in which they had edited were flagged for cleanup. It was unleashed without approval by the BAG, and some called for the “spam bot” to be opt-in only. However, PockingtonDan argued that the bot would not be useful without
sending unsolicited messages. In response, Rich Farmbrough suggested the same opt-out solution that had settled the HagermanBot controversy. However, seeing a need for extending this functionality to all possible bots, he created a template called {{nobots}}, which was to perform the same function as HagermanBot’s exclusion tag, except apply to all compliant bots. As previously described, templates allow users to insert pre-formed messages using specialized codes: the code appears when editing the page, but the message appears when viewing it. The message attached to the nobots template was blank, so it would not change the page for viewers but could be detected by bots that downloaded its source code. In this way, Farmbrough attempted to mediate between those who felt that the bot was fine as is and those who wanted to make it based on an opt-in list. In other words, if a user placed the text {{nobots}} on their user page, any bot that supported the standard would not edit that page in any fashion. A user could also allow only specific bots access by writing, for example, {{nobots|allow=HagermanBot}}. In short, {{nobots}} was a sign that users could place on pages to signal to certain bots that they were either welcome or not welcome to edit on that page, with no actual technical ability to restrict non-compliant bots from editing. A bot would have to be built such that it looked for this template and respected it; in the case of PockingtonBot, incorporating this feature would be required by the BAG in order to approve the bot:
While the controversy of PocKleanBot was settled by PockingtonDan bowing to the pressure of the bot and removing it from operation, the template fared much better in the bot development community. Along with Farmbrough, Hagerman was one of the key actors in developing the initial specification for `{{nobots}}`, along with Ram-Man, a member of the Bot Approval Group. On 18 December, Hagerman announced that HagermanBot was now “nobots aware” on the template’s talk page, the first recorded bot to become what would later be called exclusion compliant – a term that

Er, I didn't say it wouldn't be useful; I just said it should be opt-in. People don't like receiving automated messages without explicitly asking for them. I think it's a great idea and a great bot. TenOfAllTrades(talk) 20:31, 12 December 2006 (UTC)

OK I've just created a (currently null) template called nobots. Suggestion is you put it on your user/talk page and no bot will touch it. Obviously needs consensus from bot operators. Comments? (see also Template talk:nobots Rich Farmbrough, 21:00 12 December 2006 (GMT).

An opt-in method would be better. For one thing, a given editor might want to receive messages from some bots but not from others. For another, editors shouldn't be obliged to refuse spam; we should assume that people don't want spam until told otherwise. TenOfAllTrades(talk) 21:28, 12 December 2006 (UTC)

Please mark this bot as withdrawn, will not be developing this further, moving on to less controversial work! - PocklingtonDan 22:57, 12 December 2006 (UTC)

That's too bad. I hope you come back to it at some point; with an opt-in system it would be a very helpful tool, and I think that a lot editors would like it very much. TenOfAllTrades(talk) 23:01, 12 December 2006 (UTC)
Hagerman himself crafted. After some confusion with semantics, the template was copied to {{bots}} and remained relatively stable for the next few months as it gained acceptance and increased use among bots. After HagermanBot, the next bot to be made exclusion compliant was AzaBot, which was created to leave user talk page messages for users who had participated in a certain specialized discussion when an outcome was reached. AzaToth submitted the proposal to the Bot Approval Group on 20 December, which was approved by Ram-Man that same day. In his decision, Ram-Man asked AzaToth to make the bot comply with {{bots}}, implementing an opt-out mechanism to “respect their wishes.” Ram-Man also asked for AzaToth to share the source code that made this mechanism possible:

Speedily Approved. This is a proven bot from a proven user. There is no reason not to give this official approval. If you could, please make the bot compliant with Template:bots so that users can easily opt-out if they choose to do so. It seems apparent that those voting users have implicitly opted-in in this case, but you should respect their wishes if they use this new system now or in the future. Also, since other users use python frequently, any published code that does the opt-in/opt-out check would be appreciated. -- RM 13:15, 20 December 2006 (UTC)

AzaToth quickly wrote a seventy-five line function in the programming language Python that incorporated compliance with this new standard, publishing it to the bot development community. This soon became fine-tuned and reduced to a four-line snippet of code, ported to five different programming languages such that nearly any bot operator could copy and paste it into their bot’s code to achieve exclusion compliance. As members of the bot development community created software frameworks to facilitate
bot programming (including programs like AutoWikiBrowser, which allowed users to semi-automatically perform repetitive tasks), this code was eventually incorporated and enabled by default. Through the efforts of those in the Bot Approval Group and the bot operator community – especially Farmborough, Hagerman, and Ram-Man – exclusion compliance became a requirement for many bots, implemented first to settle existing controversies and eventually becoming a pre-emptive mechanism for inhibiting conflict between bot editors and the community. While it was never mandatory, many bot operators had to argue why their bot should not be required to implement such features upon review by the BAG, and failure to implement exclusion compliance or opt-out lists soon became non-negotiable grounds for denying some bot requests:

Yet {{bots}} lacked technical force on its own: as a standard, it relied on other actors to deploy it in the name of exclusion compliance or opting-out. In addition, {{bots}} was not intended to be applicable to all bots, as early discussions by bot operators indicate. Certain tasks, such as vandalism reverting and notification of copyright infringement, were deemed by their operators and the community to be so important that users should not be able to opt-out. A few semi-formal proposals were
made to amend the Bots policy to require that all bots be subject to exclusion compliance, but they failed to gain consensus. One reason was that many objectors used opt-out lists as an example of a better way that bots could be controlled. In addition, there was a strong view that subjecting all bots to such requirements would lead to disaster. Therefore, the generalized expectation in the bot development community was: bots that performed necessary tasks did not have to respect the idiosyncrasies of editors, although bots which performed less obligatory tasks ought to require more deference. Therefore, while there was a formalized mechanism for respecting an editor’s wishes when this was deemed necessary, there was no clear distinction that could be used to determine when the wishes of editors ought to be respected in such a manner. This was instead a determination that was made by the BAG on a case-by-case basis, largely requiring it when there was either an existing or potential controversy regarding the necessity of the task.

Debates about newsletter delivery bots – which exploded in popularity as the various editorial sub-communities organized in 2007 – became a site of articulation regarding this issue. Many bots were proposed that would automatically deliver a group’s newsletter or targeted message to all its members. When the first of these bots began operating, conflicts initially emerged between editors who felt they had received unsolicited spam and bot operators who thought they were providing a valuable service. Opt-out mechanisms were used to settle these disputes, although in many cases the bots already incorporated such features but did not make them visible to recipients. In response, a set of informal criteria was soon formed by members of the BAG to
ease these proposals. One requirement was implementation of some opt-out mechanism, either via exclusion compliance or an opt-out list; another was including information about opting-out in each newsletter delivery. Such requirements settled many controversies between editors and bot operators, and soon, all newsletter bots were not approved by the BAG until they were proven to sufficiently respect the wishes of editors who did not want interference from such bots.

Yet despite the work of the BAG in making \{\{bots\}\} and opt-out lists a de facto standard in respecting the wishes of editors to opt-out of bot actions, they were largely situated in a network that did not extend far beyond the bot development community. Requirements to inform users of a bot’s opt-out procedure was one way of extending these mechanisms to new networks, but this was only enforced for newsletter-delivery bots. There was also confusion regarding the role of opt-out lists or exclusion compliance in bots among those who knew of these features. In particular, some editors believed that exclusion compliance was mandatory, as the following exchange on the template’s discussion page shows:

<table>
<thead>
<tr>
<th>Can someone tell me why User:BetacommandBot was able to make this edit to my talkpage even though I had both {{bots|deny=BetacommandBot}} and {{nobots}} on the page? Skomorokh inline 21:39, 15 December 2007 (UTC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sorry for the late reply. If it is still of any interest to you, it is because BetacommandBot is not exclusion compliant. Any bot which is not in will ignore this template. Puchiko (Talk-email) 23:42, 2 February 2008 (UTC)</td>
</tr>
</tbody>
</table>

Realigning the Network: The Case of Betacommand’s Bots
The case of BetacommandBot illustrates the complex assemblage of actors that had to be continually aligned in order to make concepts like exclusion compliance and opt-out lists silently and efficiently remake Wikipedia according to this vision of how bots ought to act. When these actors became misaligned by the work of Betacommand and his allies, a controversy erupted that required a significant amount of work to settle— but was able to be settled nonetheless by re-deploying various constructed networks of associations to make the world common again. In all, the case of BetacommandBot’s non-compliance demonstrates the way in which bots are made possible in Wikipedia based on the production and negotiation of a common worldview that is supported by a cascading set of black boxes, and how Wikipedia’s technical and social infrastructures are in fact assemblages of human and non-human actors who must constantly work to produce order from disorder.

BetacommandBot was created to perform tasks related to images, including nominating images for deletion if they did not include certain information required by policy and removing images that violated policy from articles. The bot would notify users if an image they uploaded was removed or nominated for deletion in this manner, which resulted in some users being notified multiple times. Betacommand, who operated the bot, stated that the notification was too important to allow opt-outs, even when a growing group of users demanded that he make the bot exclusion compliant, or at least create an opt-out list. He had several key human allies in this issue, including the Bot Approval Group, which had become the obligatory passage point in determining whether a bot ought to implement an opt-out list or exclusion compliance. BetacommandBot
had been previously approved by the BAG, and the group did not express an opinion that BetacommandBot ought to act in accordance with the same moral worldview as bots like HagermanBot. In addition, Betacommand and his human allies could rely on the Bots policy to defend this claim, as the policies did not explicitly state that either exclusion compliance or opt-out lists were mandatory in all cases; they were instead characterized as good practices for many, but not all bots.

However, the opposing group of editors had allies of their own, although they were of a radically different configuration than the one in support of BetacommandBot. Editors like Locke Cole made formal and semi-formal proposals in various designated spaces, attempting to build a consensus for forcing Betacommand to implement a voluntary opt-out feature or exclusion compliance. One of the strongest of these assemblages was forged in a proposal to technically block the bot from editing until it allowed editors to request to not receive talk page notifications in the form of exclusion compliance. In response, Betacommand created an opt-out list, using it to respond to objections in the same way that Hagerman and his allies did. However, Betacommand’s opt-out list did not allow users to add themselves to the list, as he would retain exclusive control over who could gain entry. Users like Locke Cole, who Betacommand characterized as “lazy” and lacking “a valid reason for not getting messages” would not be added to the bot’s opt-out list. Additionally, he stated that users who opted out of notifications could not use ignorance as a valid argument in protesting the bot’s actions and would also “lose the right to complain about the bots themselves or the issues they raise.”
As the following excerpt from a discussion shows, participants relied heavily on these existing sociotechnical artifacts that had been initially used to settle the debate between Hagerman and Sensemaker. Like in that controversy, the bot operator and his allies initially held that he had no expectation to follow the requests of those who wished non-interference, referencing features of the policy environment to justify such a non-obligation:

Ok, the issue seems to hinge on that BCB currently is not compliant with NoBots because of abuse of the template in general led to it being overridden and it would be difficult to re-implement just on a user-talk/user basis. Could we try the Signpost approach? Could there be a page that users could add themselves to that is automatically (by which I mean all names, not those with a reason) are added to BCB's internal non-post list? MBisanz talk 03:37, 9 March 2008 (UTC)

Perhaps BetacommandBot should follow a "nobots" tag or have a signup list to prevent messages on user pages. But Betacommand will take requests to put users on an "opt out" list. This is not the most pressing issue for the community or Betacommand to solve at this time. It is not worth a block. (There are other issues that the community should insist that Betacommand resolve.) -- SWTPC6800 (talk)

As I already said, I requested exclusion on his bots talk page and was directed (by another editor) to try {{nobots}}. Obviously that did not suffice, because I continue to receive notifications. Note here that not only did he not add me to his exclusion list as you're saying he would, but that his bot also willfully ignored {{bots}}/{{nobots}}. This is an unacceptable state of affairs and needs to be remedied before the bot continues. — Locke Cole • t • c 05:32, 9 March 2008 (UTC)

Betacommand responded to such objections by claiming that:
However, there was a key difference in this debate: venue. Sensemaker and other anti-HagermanBot editors generally made their arguments in spaces that were not specifically designated as formal discussion pages – such as Hagerman’s user talk page. In addition, those who did enter official spaces made their pleas to members of Wikipedia’s bot development community, specifically the Bot Approval Group. Locke Cole allegedly tried the same strategy, but was unsuccessful in getting either Betacommand or the BAG to honor his request. The Bot Approval Group, which was embattled in parallel controversies regarding other tasks that BetacommandBot performed, did not make a formal determination regarding the opt-out issue – either because they supported Betacommand or because they were backlogged with the other controversies, depending on conflicting accusations. Either way, Locke Cole could not rely on them as an ally in forcing Betacommand to let him opt-out, even though many agreed with Cole’s assessment that previous bots had been required to implement opt-out mechanisms for similar notifications.

Next, Locke Cole attempted to compel Betacommand into submission by calling forth the full force of the policy environment. In his request to ban BetacommandBot, his...
audience shifted from the BAG and the bot developer community to Wikipedia’s administrative community, who had the technical ability to temporary block users. In this network, he was initially able to enroll as an ally the established procedure for blocking those who had violated Wikipedia’s foundational principles and policies, initiating a lengthy and visible discussion on the issue. Yet his shift in venue and audience had significant consequences: instead of convincing Betacommand to personally exempt him from bot-initiated messages or the BAG that such messages were unsolicited and unnecessary, Locke Cole had to convince a much larger audience to block BetacommandBot from editing in all cases until it fully complied with an opt-out mechanism. In this space, Locke Cole was not able to call forth a single policy or guideline to support his claim that BetacommandBot had violated fundamental rules, as the generalized expectation of opting-out was one that was only taken for granted within the bot development community. For various reasons, it had not been codified into policy or guidelines, even though many editors and administrators thought that it perhaps ought to be.

Even though Cole continued to refer to {{bots}} as a “de facto standard,” such a characterization held little strength outside of the bot development community; in fact, Betacommand and his allies were able to contest this by referring to discussions in which proposals to make exclusion compliance required by policy had been discussed and rejected. As such, they were able to successfully convince others that a block was not justified, even if Locke Cole’s request was one that ought to be required by the BAG. Not dissuaded, Locke Cole was able to deploy his objection in many other venues
in his attempt to bypass the BAG and policy. The conflict over BetacommandBot soon became a complex war between two networks of actors which had assembled in defense or opposition to BetacommandBot’s various tasks. As it escalated through Wikipedia, each side gaining allies along the way, it soon became clear that it would eventually have to be referred to Wikipedia’s Arbitration Committee.

ArbCom, as it is known, was instituted in 2004 by Wikipedia co-founder Jimmy Wales to adjudicate the growing number of disputes which were unable to be mutually resolved through various processes of mediation. As he wrote in his announcement of the ArbCom, it “can impose a solution that I'll consider to be binding,” and in April 2007 he stated that the Committee is justified in “overturning any decision I have made in my traditional capacity within Wikipedia.” The members of the Committee serve for three year terms, and are officially selected by Wales. Since June 2004, elections have been held to determine the level of approval for candidates, and Wales has always appointed the candidates who perform best in the election. ArbCom is taken by many to be a court of last resort for the resolution of disputes which have exhausted all other venues, although the ArbCom does not take itself to be a court. As of the Betacommand case, the Arbitration policy stated that the Committee was to:

- construct a consensus opinion made out of Principles (general statements about policy), Findings of Fact (findings specific to the case), Remedies (binding Decrees on what should be done), and Enforcements (conditional Decrees on what can further be done if the terms are met).

Members of the ArbCom generally do not enforce aspects of their decisions, instead relying on administrators and editors for this task. Its specific “remedies” are generally
framed as admonishments, encouragements, or warnings, holding that a particular editor wrongly or rightly acted in the case at hand. While some decisions include bans or blocks for wrongful editors, ArbCom rulings more often identify which actions ought to be prohibited, giving administrators the authority to enforce such compliance. As a principle, the Committee does not take cases where the content of an article or a policy is disputed, but rather when it is unclear what is required according to established policy, procedure, and principles. As such, this was a prime case for the Committee, who made the following “finding of fact” regarding the case:

In March 2008, after being asked to enable an "opt-out" feature for BetacommandBot, Betacommand created User:BetacommandBot/Opt-out, a page on which users could elect not to receive image-tagging notifications from BetacommandBot. However, in addition to making the legitimate point that users might suffer deletion of their images without prior notification if they opted out of receiving notifications, Betacommand also attempted to impose the unreasonable quid pro quo that users signing the opt-out list "also lose the right to complain about the bots themselves or the issues they raise." He failed to respond reasonably to widespread criticism that this requirement could not be justified.

In its decision, the Committee “urged [Betacommand] to establish an ‘opt-out’ list for BetacommandBot without imposing any unnecessary conditions on the right to decline to receive notifications.” While the finding was specific to Betacommand’s case, ArbCom also ruled that bot operators were, in general, subject to the consensus of the community in ways beyond what the Bots policy requires, and as such:
What is most notable about this ruling is that the ArbCom recast the issue as one based entirely in rights: editors not only had a right to non-interference from bots that could only be violated in certain cases, but bot operators also had a right to operate their programs which could also be violated under certain conditions. Interestingly, this kind of rights-based argumentation proved to be insufficient to settle the debate over HagermanBot – Hagerman was deemed to have every right to operate his bot without an opt-out list, but risked spreading botophobia and generating discord. Furthermore, in the various debates about BetacommandBot, rights-based arguments were made from all sides but failed to settle the controversy. Yet when the Arbitration Committee enrolled this previously-unformalized right as an always-existent ally in their ruling, it instantly gained force and stability, settling the issue and re-aligning the actors in relation to BetacommandBot.

Betacommand continued to operate his bots in violation of this editorial right, which resulted in the formal restriction and eventual revocation of his bot building privileges in May 2008. This social ban took the form of a restriction on unapproved automated edits, defined as a “pattern of edits” that affected more than 25 pages. This was violated by Betacommand twice: once in August and once in November, both times resulting in administrators placing temporary technical blocks his account for no more

Bot operators are expected to respond reasonably to questions or concerns about the operation of their bot. An editor who (even in good faith) misuses automated editing tools such as bots and scripts, or fails to respond appropriately to concerns from the community about their use over a period of time, may lose the privilege of using such tools or may have such privilege restricted.
than 72 hours. In December 2008, he was blocked again for automated editing, and the temporary block was later extended indefinitely after several lengthy community discussions. In January 2009, Betacommand attempted to appeal his status, as banned users are allowed to do, but after a discussion it was deemed that there existed no consensus to overturn his ban.

**Analysis**

**The Big Leviathan**

The cases of HagermanBot and Betacommand seem to be excellent evidence for the existence of a social order in Wikipedia, as they both follow the traditional genre of social narratives about technology. Both cases tell of individuals who had found ways to make themselves physically stronger than almost all others, able to singlehandedly use force to impose their view of how the world ought to be. Yet because of the social order, the will of these brutes could be checked by individuals who had far less physical power and prowess. Had the social order not existed, the strong – which in Wikipedia translates to those who could edit at the highest rate – would have reigned supreme over all others, or worse, instigated a devastating civil war as the community attempted to resist this unilateral act with only their bare hands. The social order in which all participants were embedded can thus be seen as a power structure through which the will of otherwise powerless people can be concentrated in a single actor. This Leviathan, it may be theorized, acts with the will of the entire body politic, making possible the regulation of individuals who, by virtue of their physical or technological power, would otherwise create chaos and strife in a state of nature. Constituted in an interlocking set of social
and technical sub-systems, like the Bot Approval Group, the administrative class, the Arbitration Committee, and the policy environment, order is maintained and anarchism foreclosed.

It seems relatively unproblematic to claim that the cases of HagermanBot and Betacommand occurred as they did because of the stability, force, and specific configuration of the sociotechnical infrastructure in which the participants were embedded. However, such a conclusion ignores the fact that this structure only has stability, force, and alignment to the extent to which human and non-human actors actively work to stabilize, enforce, and align it. This becomes problematic when two contradictory claims emerge out of such a conception of social and technical infrastructure. First, the resolution of a controversy in Wikipedia is taken as a consequence of the project’s sociotechnical order. Otherwise put, order is produced out of disorder because these social and technical infrastructures have stability and force and are constituted in certain ways. Yet on the other hand, it is only through the resolution of controversies that such orders and infrastructures are stabilized and enforced at all. In addition, the social and technical order shifts, that is, realigns or becomes reconstituted in one formation as opposed to another, only because the participants agree to re-conceptualize their world in relative agreement, thus settling the controversy. In short, the stability and force of Wikipedia’s sociotechnical infrastructures as well as their specific configurations explain how controversies become settled in certain ways, while at the same time, the settlement of controversies in Wikipedia is what gives these infrastructures stability and force as well as explains their specific configuration.
This paradox is the same one that Bruno Latour has identified in his study of scientific practice between what he calls “science in action” and “ready made science” (Bruno Latour 1987). For him, the former refers to scientific facts which are not yet settled or agreed-upon and the latter refers to scientific facts which are generally accepted as true reflections of Nature. After a scientific controversy is settled, it becomes clear that the victor was correct because he or she was right; Nature or the truth was in his or her corner all along. However, when a controversy is still active, all participants claim to have the truth of Nature on their side, meaning that something other than Nature has to settle the controversy:

If you wish to attack a colleague’s claim, criticize a world-view, modalise a statement, you cannot just say that Nature is with you; ‘just’ will never be enough. You are bound to use other allies besides Nature. If you succeed, then Nature will be enough and all the other allies and resources will be made redundant. (98).

Because of this, Latour’s accounts of scientific controversies show how one participant is exclusively able claim that Nature was on his or her side all along. In such investigations, the pre-existing stability or externality of Nature cannot be taken for granted and used as a reason why one fact came to be accepted as true instead of another. This position, which Latour calls relativism, must be taken because it is how the participants themselves act when arguing over facts:

When studying controversy … we cannot be less relativist than the very scientists and engineers we accompany; they do not use Nature as the external referee … since the settlement of a controversy is the cause of Nature’s representation not the consequence, we can never use the outcome – Nature – to explain how and why a controversy has been settled. (99).

At this point, it may seem that Latour wants to give a social explanation of
scientific facts, that is, to show that ‘mere’ social factors are the reasons why participants are made to accept or reject a proposed finding. For example, one fact becomes victorious over another not because it actually reflects the truth of Nature, but instead because its spokesperson was able to use rhetoric, coercion, economics, bureaucracy, ideology, and so on to compel others to take it to be true – independent of its actual truth-value. Because the participants are embedded in a social structure, it may be assumed that the victor is the individual who is best able to navigate or manipulate it in order to force others to accept the validity of their fact over others or face dire consequences. A slightly different view might put less emphasis on individuals and more on structural conditions: the scientist with the biggest lab, most funding, greatest prestige, and so on, is categorically the one whose facts are accepted as true. In all, these explanations are generally based on groups who have motivations for accepting or rejecting a fact, and that the study of science or technology should critically examine the interests of such groups in determining whether a fact is accepted or rejected.

While many sociologists of science have taken this route, Latour contests such efforts and refuses to claim that such a social view of scientific practice necessarily demonstrates the falsity of scientific facts. He contrasts his descriptions of science with traditional social explanations of religion, which may demonstrate the falsity of a divine law by showing how it was followed because of coercion and force instead of devotion and faith (Bruno Latour 2000). Instead of replacing scientific explanations with social explanations, Latour’s study of science holds that scientific facts gain their truth, objectivity, and congruence to Nature precisely because of the work done by
these human and non-human actors. In an earlier study of laboratory practice in which a fact about a molecule called TRF was constructed, he and Steve Woolgar state: “to say that TRF is constructed is not to deny its solidity as a fact. Rather, it is to emphasize how, where, and why it was created.” (Bruno Latour and Woolgar 1986, 127). As such, Latour rejects the social determinism that seems necessary once Nature has been re-defined as that which is produced by the settlement of scientific controversies, arguing that:

Analysts who use groups endowed with interests in order to explain how an idea spreads, a theory is accepted, or a machine rejected, are not aware that the very groups, the very interests that they use as causes in their explanations are the consequence of an artificial extraction and purification of a handful of links from these ideas, theories, or machines (Bruno Latour 1987, 141).

In its place, Latour proposes a view of society that mirrors his previously-established view of Nature:

Once the controversies have ended, then a stable state of Society, together with a stable rendering of the interests of its members, will emerge … Not so, when we follow facts in the making. … since the settlement of a controversy is the cause of Society’s stability, we cannot use Society to explain how and why a controversy has been settled. (143-4).

While observers of scientific practice can forge deterministic explanations of its order based on Nature or Society, this is only the case because scientific controversies are resolved by the construction of various black boxes and re-alignment of various human and non-human actors whose existence and form we now cannot otherwise conceptualize. Yet before controversies are settled, the black boxes were open and the actors misaligned; in fact, this divergence did not spark the debate, but is often the core issue that had to be resolved.

Steven Shapin and Simon Schaeffer’s description of scientific
practice in Europe during the Protestant Revolution (1985) shows a case that perfectly exhibits this phenomenon. As the authors show in their analysis of debates between Thomas Hobbes and Robert Boyle over the nature of vacuums, the conflict is predominantly cast as one between an obviously correct scientific position supported by evidence and an obviously wrong philosophical position masquerading as science; yet such a view is only possible because of the way in which the debate was settled. As they argue, the debate between Boyle and Hobbes was not merely one about natural facts, such as if a purely empty vacuum was possible. In addition, the two adversaries differed in their methods of knowledge production to such an extent that we take Boyle’s claims to be based on the first modern science experiments, while we hardly recognize Hobbes’s arguments – which were based in the long-established methods of natural philosophy – as valid on their own. While Hobbes crafted logical arguments and thought experiments in order to make claims about Nature, Boyle built a device which would make natural phenomena visible and then demonstrated it in a quasi-public space. While we instantly recognize Boyle’s methods as scientific, he had to first convince his peers and adversaries that his method could be indeed used to make claims on its own, that is, without needing the kinds of a priori philosophical reasoning seen as necessary in natural philosophical circles. More importantly, he had to convince others that such a mode of knowledge production could, on its own, discredit seemingly-obvious and logically airtight philosophical thought experiments without providing a philosophical counter-explanation.

After Boyle had convinced almost all others that experimental science ought
to trump philosophical reasoning when the two collided, the only confusing element of the Boyle/Hobbes debates is the extent to which the public clung to philosophy in the face of contradictory empirical evidence. This is a position we often attribute to mere irrationality, as it does not seem that we need to explain why Boyle’s experimentation replaced Hobbes’s philosophy in the realm of Nature. However, as the authors transport us back to the time of the controversy, it becomes clear that Boyle faced steep challenges in making it ‘obvious’ that his claims ought to be accepted instead of Hobbes’s. Shapin and Schaeffer argue that it was only through the creation of various technologies – social, literary, and material – that Boyle was able to transform the activity of natural philosophy into modern science, relying more on scientific experiments than thought experiments. The elements of experimental science that we now take for granted were not yet settled, and their absence or malfunction posed significant challenges for Boyle and his allies.

For example, without esteemed scientific bodies like the Royal Society, who would witness experiments and issue authoritative reports, it was impossible for an experimenter’s claims to travel with as much ease and speed as those made by natural philosophers. Likewise, without the genre of formalized laboratory reports, which went into significant detail regarding the construction of experimental apparatuses, independent replication of experiments and verification of findings was equally impossible. However, after these and other technologies had been developed and deployed, the claims of experimental scientists like Boyle were able to become pervasive across Europe. In fact, the authors claim that the Boyle/Hobbes debate about the nature of vacuums was resolved not because Boyle showed that Nature had been on his
side all along, but because of the construction of such technologies that we now take for
granted. In this light, Hobbes’s objections can be seen as more than outdated
irrationality: without the possibility for the independent replication of experiments, would we moderns be entirely justified in taking experimental science to be an entirely valid
mode of knowledge production?

**Actualizing Consensus**

The case of HagermanBot is similar to the Boyle/Hobbes debate, as the
development and deployment of various technologies (broadly defined) made possible a
resolution to the debate that may not have otherwise resulted. In the same way that
laboratory reports made possible activities like independent verification and convinced
many that experimental science was a credible method of knowledge production, so did
the technology of opt-out lists and exclusion compliance make possible the right of non-
interference from bots. This actualized right, articulated by Sensemaker as a plea for
mutual respect and by Farmbrough as a “bots are better than people” mentality,
convinced the participants that a continuously-operating bot could enforce a guideline
while respecting editors and therefore preventing botophobia. At present, it seems
unthinkable that a Wikipedian bot would be authorized to unnecessarily transform a
user’s comments or user talk pages without providing an opt-out list or enabling
exclusion compliance – which is why Betacommand faced expulsion from the
community for repeatedly violating this fundamental principle.

At the time of the HagermanBot debate, neither the technology nor normative
backing of opt-out lists had been developed, and was there was little more than an unsettled expectation that users had general but not total autonomy over how other editors could modify their user pages and comments. Most importantly, the concept of rights as they applied to bot actions was not officially articulated, but claimed by all sides to support their position – both before and after the HagermanBot controversy. A few bots had idiosyncratic exclusion mechanisms such that editors could direct them to not edit pages when their algorithms were malfunctioning – which Hagerman had implemented in his bot almost from the beginning. Yet as previously stated, these opt-out mechanisms were not premised on the rights of editors to opt-out of otherwise valid bot actions that they opposed for personal or irrational (e.g. counter-normative) reasons. HagermanBot’s opt-out list was different in that it allowed users like Sensemaker to decide for themselves if a guideline like Signatures ought to apply to them, and the criteria for making such a decision was entirely their own. Again, as Betacommand’s ban shows, it was soon obvious that bot operators were not allowed to control access to opt-out lists or impose ‘unreasonable’ criteria in order to gain access.

At the time of the controversy, Sensemaker was adamant that he did not need to present a rationale for deviating from the Signatures guideline in order to gain entry on an opt-out list, and in the Wikipedian community, such a position seems retrospectively rational. Given that editors are generally taken to have a right of non-interference from unnecessary bot actions, it seems to be unsurprising that Sensemaker won out. Yet despite this contemporary obviousness, it is critical to understand that this right is only recognizable at all – much less as one enabled by opt-out mechanisms – because
a significant amount of work was performed to articulate, develop, legitimize, codify, and blackbox this right such that it could eventually be taken for granted and deployed across various heterogeneous networks. The outcome of the debate should also not be taken for granted: editors like Fyslee had many allies to call forth in only opposing Sensemaker’s request, able to characterize it as a “deviation” often indicative of malevolent intent. As such, the right of non-interference from bots had not yet existed in any codified or explicit form prior to the HagermanBot controversy, and did not need to come into existence at all. Because of this, Sensemaker could only call forth a seemingly weak ally against the assemblage of actors aligned against him: an injunction for mutual respect based on “a gentleman's sense of decency and reciprocity,” which was far outside of the scope of Wikipedia’s policy environment and was characterized as a mere mask for malevolence.

Yet when Rich Farmborough entered the debate and advanced the “bots are better than people” maxim for bot operation in support of Sensemaker, it instantly reconfigured the network of actors present in the debate. He acknowledged that Hagerman and Fyslee were indeed correct in their interpretation of what the policy environment allowed, but enrolled a non-policy actor to support implementing an opt-out list anyway: botophobia. As such, Sensemaker’s plea for gentlemanly respect was transformed from either a quaint gesture or a mask for malevolence to a strong argument, albeit not on its own terms. Even though HagermanBot had every right to automatically sign all unsigned comments as established by the policy environment, Farmborough’s view of how bots ought to be developed held that a risk of increasing tensions between anti-bot editors and bot
operators was not worth whatever gain would be provided by the universal enforcement of a guideline. Hagerman agreed and implemented the opt-out list, but it would be premature to state that at this instant, the right of non-interference from bot actions came into being. All that the participants agreed upon was the use of an opt-out list on the grounds that it would settle the controversy to the satisfaction of all participation. Even the rationale as to why this particular compromise was desirable shared no unified theme: for Sensemaker, it was the realization of an ethos of mutual respect; for Hagerman, it was a consequence of making bots behave better than people; for Farmborough, it was a way to prevent botophobia. It is of note that none of these rationalizations were framed in terms of rights; in fact, Hagerman was quite explicitly deemed to have a right to ignore Sensemaker’s pleas. Instead, editorial rights were trumped in favor of settling the controversy.

It was only through a significant amount of work performed by humans and non-humans that we can now look back and frame the debate as one over rights of non-interference from bot actions. In particular, the work of the BAG in requiring opt-out lists or exclusion compliance for many bots was critical in naturalizing this requirement, turning it into an expectation for bot operation that often did not even have to be explicitly enforced by the BAG. Most interesting is the fact that Farmborough’s philosophy that “bots are better than people” – which Hagerman took up with zeal in developing standards for exclusion compliance – was left behind in this translation, as was the fear of increased botophobia from editors. Even though this seemed to be the key higher-order rationale for the resolution of the controversy between
Hagerman and Sensemaker, this view failed to take hold when the principles which emerged from the settlement began to be formalized and naturalized. When opt-out lists and exclusion compliance as technical artifacts began to be extended into the network of actors that constituted the bot development community, it lost such a normative foundation. Instead, opt-out lists and exclusion compliance simply became one of many taken for granted elements of what it means to be a bot, supported by various actors who were aligned with each other in often seemingly-contradictory ways. In fact, such decoupling was a result of this extension: these technological artifacts were able to be forcefully deployed in subsequent discussions regarding bot development precisely because they were justified on non-normative grounds, making possible a compromise when arguments based on rights or policy proved difficult to adjudicate.

**Circulating Blackboxes**

The concept of opting-out was unproblematically blackboxed within a single network of associations, one that was primarily constituted around the Bot Approval Group and focused on regulating bots. While this network was a complex and heterogeneous assemblage of human and non-human actors whose relationships and practices were often hotly contested, it was itself blackboxed in Wikipedia: for many, especially in the administrative community, the BAG was fed issues regarding bots and transformed them into official stances using their technical and normative knowledge. For a non-BAG actor, little technical, institutional, or normative knowledge was necessary in order to know, for example, if a certain bot was authorized to operate in a certain way. When the concept of opting-out and its associated software
mechanisms were deployed inside the bureaucratic black box of the BAG, it worked rather well to settle controversies. Soon, its use stabilized as it became a taken for granted element of the bot approval process within this assemblage, deployed preemptively to the extent that it became an internal de facto standard. However, this concept had little ability to circulate within other networks, particularly those involved with administration.

The controversy over Betacommand was unique due to its relation to these assemblages: having been approved by the BAG as-is and unable to get a pronouncement from them on the issue, Locke Cole and his allies found themselves unable to extend the black box of opting-out into new spaces. While Locke Cole was in one sense correct to claim that it was a “de facto standard,” this was only the case in the bot development community, whose networks would have to be extended in order for it to apply in administrative decisions. As such, when attempts to convince first Betacommand and then the BAG failed, opting-out had to be extended into new networks that initially proved problematic, as there had been no translation of expectations of bot behavior enforced by the BAG into official policy. In requesting that the bot be banned for non-compliance of the standard, Betacommand and his allies were able to deploy policy in support of their position, while Cole and his allies were only able to express frustration that was empathetic, but not actionable.

In debates within the Wikipedian bot development community, it is clear that opting-out had become a de facto standard, backed up by a normative expectation whose ubiquity was a product of a significant amount of work over years. Some bots
fulfilled this requirement using exclusion compliance, others used opt-out lists, and many more did not need either due to their actions or scope. While some bots performed actions which were taken to be necessary and therefore did not incorporate an opt-out mechanism, nearly all bots operated under the premise that editors ought to be able to avoid unnecessary interference from bot actions. This taken for granted element of bot development was largely uncontestable within the bot development community, although there were still debates about the necessity of various bot actions in light of this generalized expectation. However, it is critical to understand that the debate about BetacommandBot was not over the necessity of the notifications it sent, even though this was the initial point of disagreement between the two groups.

The BAG and Betacommand took the bot’s notifications to be necessary, and because of this no opt-out mechanism was required for BetacommandBot – although one was voluntarily implemented to not notify editors who had ‘valid’ reasons. In contrast, Locke Cole and his allies took the bot’s notifications to be unnecessary spam, arguing in favor of an opt-out mechanism that would allow any editor to decide for themselves if they should receive such messages. The problem was not that the BAG was, as some claimed, a cabalistic regime, but that a certain expectation had been developed within this blackboxed assemblage which was difficult to extend to another network. When Cole claimed that the BAG had failed to enforce the opting-out standard on BetacommandBot, he was prepared to show that the bot made unnecessary actions and therefore ought to be required to implement opting-out.

Yet this was not enough to sustain the concept of opting-out and its durable
mechanisms in administrative networks. Cole was prepared to perform the argumentative work required to convince members of the BAG that they ought to force BetacommandBot to adopt opting-out. For whatever reason, he failed to do so within the bot development community, which was embattled in what were seen as more serious controversies regarding Betacommand. Then, Cole attempted to use the same strategy to convince administrators that they were obligated to perform the same action: force BetacommandBot to comply with all opt-out requests or face a block. Yet precisely because of the blackboxing work required to establish opting-out as a de facto standard within the bot development network, opting-out was not recognized as an actionable policy requirement in the administrative network, and its mechanisms were not seen as enforceable standards. While many sympathized with Locke Cole and felt that BetacommandBot ought to be forced to implement an opt-out list without restrictions or even that all bots ought to be subject to such a requirement, as one user curtly stated: “Should be, but, is not. WT:BOTS is the second door on the left,” referring to the talk page for Wikipedia’s Bots policy. However, discussions in that space proved to be just as futile as they were when they were perennially proposed, as an abstract criteria for determining the conditions under which a bot ought to be subjected to such requirements was difficult to formalize in such a way that satisfied all – or even most – participants.

When the Arbitration Committee reviewed the case, they settled the controversy by introducing a new actor: the rights of editors to not be subjected to unnecessary interference from bot actions. In the same way that introducing opting-out into the HagermanBot controversy was a way of reconfiguring the associations of the actors
involved such that a settlement was made possible, so did the ArbCom ruling perform a similar task. These formalized standards created by ArbCom altered the way that opting-out could be deployed, making possible a sanction of Betacommand by the administrative assemblage. However, the crucial difference was that ArbCom’s conception of these rights was folded into what it meant to be a bot editor, including the injunction to respond to community concerns with civility and respect. As such, Betacommand’s bot operating privileges could be formally revoked by administrators when he violated not the de facto standard of opting-out, but instead the expectations of a bot operator. This occurrence should not be taken as a purely juridico-political action, but instead as a translation of the blackboxed concept of opting-out from the network of bot operation to the network of administration.

**Conclusion**

In all, the case of HagermanBot shows us not how a weak but pre-existing social norm was controversially reified into a technological actor, but rather how a controversy regarding the delegation of work to non-humans was settled by the construction of a new set of technical and social artifacts, which the Wikipedian bot development community used to reassemble their social and technical order. After the Betacommand controversy, we can look back at the HagermanBot controversy and easily claim that the participants were actually in the process of formalizing an individual right of non-interference from bot actions, even though the settlement was cast in terms of botophobia and social cohesion. While this is most certainly a valid argument, it is only the case because of the work performed in settling the Betacommand controversy on the basis of
individualistic editorial rights. As such, the case of BetacommandBot complicates the depiction of opt-out lists and exclusion compliance as taken for granted elements which invisibly actualized a right of non-interference from bot actions from their inception. The traditional sociotechnical determinist narrative would claim that these artifacts, once constructed, sunk into the Wikipedian lifeworld and invisibly manipulated the actions of humans in spite of themselves. Instead, this case shows that opt-out lists and exclusion compliance as either necessary requirements for bot operation or as sociotechnical reifications of an abstract editorial right can only be black boxes that silently manipulate human action when there is an aligned network of actors that unproblematically deploys them in such a manner. When these black boxes are pushed into unfamiliar networks, they lose their taken for granted nature and must be translated into such a form that allows their uptake within such new assemblages.

This chapter has contested accounts of the project’s order that rely almost exclusively on social artifacts, showing that these non-human editors have a significant effect on the way in which the project’s norms are enforced. While much human work (facilitated by social artifacts) is performed in the settling of controversies, the bot development process can be a moment of articulation for what were previously taken to be uncontroversial expectations. In contrast to accounts of Wikipedia’s order that focus almost exclusively on social artifacts and interactions, the negotiation of a bot’s source code is not a purely normative affair in which participants discuss the kind of editorial environment that is to be enforced by such an actor. As the case of HagermanBot shows, the construction of new technological actors commonly aligned with competing
normative expectations can help make the wiki world common again. Seemingly insignificant concepts like exclusion compliance or opt-out lists, having become black boxed and normalized, were shown to hold together a common understanding of how bots were to operate in Wikipedia. When this was contested, as in the case of BetacommandBot, the dispute was likewise not a purely social affair between human actors in a pre-existing social order disputing the desirability of a technology. In this moment of controversy, associations between humans and non-humans became increasingly problematic as what was claimed to be a de facto standard failed to translate into this network. The work of making the social world common again in the context of this bot required the negotiated articulation of a technical concept like exclusion compliance or opt-out lists; as shown in the ArbCom decision, translating this into an editorial right was critical in settling the controversy.

These cases show that bots have much stronger role in the explaining Wikipedia’s order than the previously-theorized force-multipliers of power users. Instead, bots like HagermanBot actively work to realize a highly-particularized vision of how encyclopedic knowledge production ought to occur, including such seemingly-trivial norms like signatures for comments in discussion spaces. Such accounts also contest the “wisdom of crowds” model of autonomous group decision-making as an explanatory theory for Wikipedia’s order. First, the mere existence of bots that enforce norms shows that a significant portion of the project’s social structure does not emerged pre-formed out of the multiplicity of humanity – as its critics and supporters both tend to allege. Instead, it must be constantly produced by actors like HagermanBot and BetacommandBot, who
actively and unceasingly work to realize a particular normative vision of wiki-based encyclopedia building. Second, these case studies demonstrate that the process of dispute resolution is far more than a simple economic aggregation of interests and opinions, that is, a market-based task which would, for followers of Surowicki, be ideally performed in isolation from others. This view of Wikipedia is unable to account for the way in which a shift in associative networks (from bot developers to administrators) had a transformative effect on the way in which both a technological standard and a normative expectation could be made comprehensible. While the roles of such technological actors in transforming the average, everyday activities of editing and administration remain unarticulated, this chapter has demonstrated their importance in working to produce, maintain, and contest Wikipedia’s social order.
Chapter 3. Vandal Fighting

Introduction

This chapter examines the process of vandal fighting in Wikipedia, examining the way in which editors and administrators review contributions to Wikipedia and enforce various normative and epistemological standards on infringing edits. Scholarly and popular accounts of the project often explain Wikipedia’s near-immunity to vandals, spam, and ideologues by positing a staggering number of insomniac reviewers; the project’s administrators, for example, must be constantly at work in seeking out and blocking malevolent users in order to keep the project from degenerating into anarchy. While such a view is partially correct, it largely ignores the heterogeneous assemblage of human and non-human actors who are deployed in the routine identification and blocking of a vandal. This chapter presents one such case, focusing on an anonymous user who was temporarily blocked from editing after making approximately twenty inappropriate edits to various articles during a one hour period. Almost all of the edits made were identified as vandalism and reverted by a diverse group of users using many different tools to interact with Wikipedia.

Such technological actors are shown to actively reshape the way in which editors experience their project and edits to it, making possible a kind of epistemological enforcement that requires little to no specific knowledge about the article at hand. In addition, this account shows that the process of vandal fighting is not a disconnected
activity in which atomistic editors enforce their view of the world on others, assisted by technological force-multipliers. Vandal fighting is instead shown to be a process of distributed cognition, through which users can come to know their project and the users who edit it in a way that would be otherwise impossible for a single individual. Taking from the work of Edwin Hutchins, this chapter claims that in same way that the navigator of a ship can know trajectories only through the work of dozens of crew members, so is the blocking of a vandal a cognitive process made possible by a complex network of interactions between humans, encyclopedia articles, software systems, and databases.

The hour between 21:00 and 22:00 (EST) on Thursday, 19 February 2009 was a rather routine one in Wikipedia. It was one of the busiest hours that day, but this is expected, as contributions to Wikipedia on weekdays generally peak during primetime on the East Coast of the United States. In this hour, about 12,738 edits were made – an average of 3.5 per second – to 8,099 distinct pages in the wiki. These contributions were made by 3,999 unique users, and about half of these users were unregistered and anonymous, identified only by their IP address. However, anonymous users only made about one quarter of all edits to Wikipedia during this time. About 15% of all edits were made by bots, while administrators contributed about half that amount. Non-administrative, non-bot registered users – around 1,784 in all – made about half of all edits in the hour. During this time, nineteen distinct administrators also temporarily blocked a total of forty-one users or IP address from editing and temporarily protected eleven pages from editing. In all, such frequencies are roughly representative of a weekday primetime hour in Wikipedia.
Table 1: Editing frequencies for 21:00 and 22:00 (EST), 19 February 2009

<table>
<thead>
<tr>
<th>User type</th>
<th>Total edits</th>
<th>Unique users</th>
<th>Edits per user</th>
<th>Edits per minute</th>
<th>Edits per second</th>
<th>Percent of all edits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anonymous</td>
<td>3424</td>
<td>1897</td>
<td>1.80</td>
<td>57.07</td>
<td>0.95</td>
<td>26.88%</td>
</tr>
<tr>
<td>Bot</td>
<td>1938</td>
<td>52</td>
<td>37.27</td>
<td>32.30</td>
<td>0.54</td>
<td>15.21%</td>
</tr>
<tr>
<td>Administrator</td>
<td>969</td>
<td>166</td>
<td>5.84</td>
<td>16.15</td>
<td>0.27</td>
<td>7.61%</td>
</tr>
<tr>
<td>Reg. editor</td>
<td>6407</td>
<td>1784</td>
<td>3.59</td>
<td>106.78</td>
<td>1.78</td>
<td>50.30%</td>
</tr>
<tr>
<td>Total</td>
<td>12738</td>
<td>3899</td>
<td>3.27</td>
<td>212.30</td>
<td>3.54</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Figure 3.1: Double logarithmic plot showing weekly user editing rates
Furthermore, statistics generated from a broader time period indicate that a significant number of users are infrequent contributors to Wikipedia, making only a few edits to the project in a given week. An analysis of all contributions made to Wikipedia between 3 February and 31 March 2009 – which were gathered using a custom bot that retrieved a list of all recent changes on a regular interval – shows that over 700,000 users made an average of one edit per week or less during this time. Over 400,000 users made between one and two edits per week, about 100,000 made between two and three edits per week, and as the double logarithmic graph in Figure 1 shows, this trend continues downward.

Given such statistics, the immediate question is how such a staggering number of diverse contributions can be controlled in any meaningful manner. Common knowledge holds that many of these edits will be undesirable and even malevolently made, given that Wikipedia does not even require users to register an account in order to edit most pages. While some supporters of Wikipedia enthusiastically embrace what they perceive to be an anarchistic field of unrestrained knowledge production, far more level this same charge when critiquing the project. For critics like Goldman (2009) and Sanger (2005b), Wikipedia’s openness and consequential lack of epistemic regulation is troubling, as they claim it will inevitably lead to widespread disruption as vandals, ideologues, and other troublemakers are allowed near-unrestricted access. Interestingly, both of these critics hold that Wikipedia cannot be saved from within, that is, by implementing stronger epistemic standards or technological mechanisms that make editing a more selective process. Instead, they claim that the project’s culture is fundamentally flawed.
and, especially in the case of Sanger, has little chance of reformation. This is the primary reason that he has given for founding Citizendum, a wiki-based encyclopedia that is intended to transplant Wikipedia’s mission into a community that has, for example, a more serious commitment to knowledge production and greater deference for credentialed experts.

The problem with such claims – especially those of Goldman, who has been predicting the death of Wikipedia since 2005 – is that they unproblematically make a causal link between editorial openness and anarchism in the editorial process. Oddly, supporters of Wikipedia have generally rebutted these critiques by accepting such a premise, arguing that editorial anarchy is desirable for a number of reasons. In such defenses, the famous dictum “with enough eyes, all bugs are shallow” – authored by Linus Torvalds, the founder of the open-source Linux operating system – emerges with surprising regularity. *Linus’s Law*, as it is called, is often invoked in a mystical manner, used to defend the observation that increased participation necessarily leads to increased reliability, independent of the qualifications or knowledge of particular members. In such a view, Wikipedia’s order is only possible because there are many individuals who are constantly at work patrolling the site for vandalism, error, or bias. Critics like Goldman have seized on this rebuttal by arguing that the volunteer human labor force required for Linus’s law to take effect is unsustainable.

The problem with this defense and criticism of Wikipedia is that the underlying explanation for the project’s order is obscured as it is reduced to the spontaneously-emergent, micro-to-micro ‘wisdom of the crowds’ style explanation. Within this
frame, the activity of insulating the encyclopedia from epistemic disorder seems impossibly daunting, given that Wikipedia has millions of encyclopedia articles and that edits, at peak times, can average 3.5 per second. In response, recent social scientific accounts of Wikipedia have begun point to the project’s administrators for an explanation: for whatever reason, these users are so involved in the process that they collectively hold together the stability of the project. These elite users are compared to police, judges, village elders, or other positions of authority, and are seen as the enforcers of a social structure that does not largely concern most contributors. This group authors the rules and standards, decides on their application in times of controversy, and bans those users who are disruptive to the encyclopedia-building process. Because of this, the predominant sociologically-themed narrative of Wikipedia is that users can edit in any number of ways – including malevolently – because these administrative elites are there to ensure that good contributions are accepted and encouraged while bad contributions are removed and punished.

On the surface, this account of Wikipedia seems to be the case, particularly with the banning of problem users. Only administrators have the technical and social authority to temporarily block users from editing, and the regularity of blocks (once every minute and a half, at peak times) seems to indicate that administrators are constantly trawling through Wikipedia in order to find and remove these vandals, spammers, and ideologues. Likewise, the regularity of page protections and page deletions also serves as evidence for a hyper-active, insomniac administrative cabal, who collectively keep watch over the encyclopedia project. Yet this characterization of Wikipedia’s stability
ignores the material practices involved with the banning of a user, protecting a page, or performing any other routine administrative action required to produce order out of the disorder produced by problem users. When this process is examined in detail, it becomes clear that ‘vandal fighting,’ as it is colloquially called, is not a process performed by a small group of elites, but is instead a diverse process that involves a significant number of human and non-human actors. It is shown how various bots and assisted editing tools make possible a kind of non-encyclopedic knowledge production regarding contributors. Statements like “user [x] is a vandal and must be blocked from editing” do not emerge full-formed out of the heads of ‘power users,’ as is often implied. Instead, this chapter details the non-human actors who make possible the aggregation, analysis, and synthesis of information from various sources, which in Wikipedia is required in order to make accountable claims of this sort.

The most problematic issue with this framing of Linus’s Law when applied to Wikipedia is not its easily rebuttable claim of cabalism. Instead, it is the way in which such accounts take for granted the existence and stability of various macro-actors who can be invoked to explain Wikipedia’s order or disorder. Interestingly, this tendency is made by critics and supporters alike, as well by researchers who otherwise present conflicting views of the project. Wikipedia’s “culture” (Wray 2009), “bureaucracy” (Butler, Joyce, and Pike 2008), “activity system” (Bryant, Forte, and Bruckman 2005), “organizational structure” (Stvilia et al. 2008) and “policy environment” (Beschastnikh, Kriplean, and McDonald 2008) are some of the most frequently invoked examples among those who do not make the naïve claim that the project is a lawless anarchy.
These social structures or macro-actors are taken to invisibly guide the actions of Wikipedians in their daily tasks, be they administrative actions, conflict resolution, or encyclopedia editing. While researchers making these claims often acknowledge that this work is increasingly being delegated to non-administrators, research at this level has predominantly focused on talk page discussions of article content, specifically looking at the role of administrators as well as official policy and procedure in the negotiation of epistemic claims.

For these researchers, social elements – human actors with various social statuses, designated workflows, codified norms, dispute resolution processes, enforcement practices, and so on – are taken to coalesce into a social structure which produces order out of disorder. While this social structure may enable participatory self-governance, this view holds that they are maintained almost exclusively by work performed by a dedicated group of human actors. A perfect example of this is Forte and Bruckman’s study of policymaking and enforcement:

the Wikipedia community has remained healthy in large part due to the continued presence of “old-timers” who carry a set of social norms and organizational ideals with them into every WikiProject, committee, and local process in which they take part (2008, 10).

The authors do devote a few paragraphs to the role of technology, but like most socially-heavy accounts of Wikipedia, it focuses exclusively on the open editorial technology. Citing the work of Langdon Winner, they claim the software system upon which Wikipedia has been built enables a more participatory mode of governance, especially compared to other on-line communities:

The potential for sophisticated community-generated social norms and
governance mechanisms is partly a designed feature of the technological architecture in which the community grows. In other words, artifacts have politics (see [20]) and code influences the development of policy. In Wikipedia, we find that code, policy, and social norms all support the wide distribution of power to govern (2008, 10).

Forte and Bruckman’s analysis of the sociality of code does not extend beyond the claim that the social macro-actors (e.g. governance, social norms, policy) are constructed and enforced in a more distributed manner because of the specific configuration of a technological macro-actor. Their work also does not take into account the role of the software in enforcing social norms – for example, blocking a problematic user for excessive vandalism or automatically altering comments to bring them in line with discursive standards. Given that their focus is on the social aspects of discussion pages, this is not surprising; however, it does indicate a gap in the Wikipedian literature.

**Problematizing Macro-Actors**

Such research tends to take for granted the stability and unity of the various social and technological macro-actors that are claimed to structure human action. Homogeneous entities like ‘code,’ ‘culture,’ or ‘governance’ come on the scene to explain user behavior – and each other – despite the fact that these macro-actors only have stability and unity to the extent that humans and non-humans work to stabilize and unify them. While many researchers of Wikipedia (Stalder and Hirsh 2002; Forte and Bruckman 2008; Bryant, Forte, and Bruckman 2005) have made passing references to the effects of project’s software architecture by citing Lawrence Lessig’s famous declaration that “code is law” (Lessig 1999), this relationship is far more complicated if we open up such a black box by asking what exactly constitutes Wikipedia’s code. A default
installation of MediaWiki, the software system upon which Wikipedia runs, has over 600,000 lines of code in 900 files, each of which is connected to the other in often convoluted ways – as any software developer knows. In addition, the version of MediaWiki that is used for Wikipedia contains over forty extensions, which are modular packages that can alter nearly every aspect of the software system. Some of these provide new editorial functionality (like EasyTimeline, which enables the creation of standardized, dynamic timelines in articles), while others perform automatic filtering for certain kinds of edits (like SpamBlacklist, which lets administrators block pages from being edited if they contain a link to certain websites.

To what extent are these extensions a part of Wikipedia’s code, such that they are elements of the macro-actor that structures the nature of normative authority in Wikipedia? Such a question seems pedantic, but controversies over proposed extensions – like Flagged Revisions, which would have required that every edit be approved by a trusted user before it went live – illustrate the fragmented nature of software systems. Something as homogeneous and monolithic as ‘code’ can only come on the scene as a recognizable macro-actor capable of structuring society when a significant amount of work is performed to produce and maintain complex relationships between a highly heterogeneous group of human and non-human actors. In addition, maintenance of a software system – like a museum (cite something) – is not merely an activity of preserving pre-existing relationships between these pre-assembled actors. Instead, such work involves actively shaping the way in which code snippets, modular extensions, data formats, protocols, user interface designs, and many more non-human entities
interact with both each other and humans who are to access the system with a variety of intentions and skill sets.

Finally, the pre-existing stability and unity of code as a macro-actor can be problematized by various unofficial or semi-official software programs and code-like elements. As the previous chapter demonstrated, autonomous editing programs (bots) can have a significant role in relation to normative standards, making possible the near-universal enforcement of any editorial expectation that can be translated into an algorithm. The case of HagermanBot contested both technological and social determinist narratives when such a translation from norms to code was contested. While the bot in question was officially approved, the case of BetacommandBot presented later in that chapter showed how a bot’s official status became contested. While previous researchers have questioned the extent to which norms ought to be encoded into software processes, the issue at hand – which must be resolved before such a political debate can occur – is the extent to which such bots can be included in the macro-actor of ‘code.’ This chapter aims to open this question even further by showing how a seemingly routine social practice – the banning of a vandal by an administrator – is not held together by an already-existing social order, but instead must be constantly structured by the work of various software agents. While many non-human actors in this account are directly incorporated into the MediaWiki code, the majority of them would not be included in an orthodox definition of the project’s software architecture. Developed by members of the vandal fighting community to assist with various tasks, these non-human actors have a substantial effect, working diligently to hold together the process of normative
enforcement among human editors.

The account of vandal fighting described in this chapter is inspired by the work of Bruno Latour and Michel Callon. As they claim, the problem with a traditional conception of sociological or technological macro-actors is not that they are fictitious – Latour and Callon emphasize that they are very much real, monolithic, and in force. Rather, they interrogate the conditions of possibility for entities like ‘the state’ or ‘science,’ finding that they have such a monolithic unity because of a significant amount of work is constantly performed by both humans and non-humans to unify and stabilize them. Specifically, in their essay “Unscrewing the Big Leviathan,” (1981), the authors claim that there is no a priori difference between macro-actors and the micro-actors who they dominate or structure. Instead, states, corporations, cultures, populations, and disciplines are an achieved, constructed assemblage of many micro-actors who must be constantly aligned to unify the macro-actor. The core of their claim is in reconceptualizing macro-actors as an effect that must be explained, not a cause that explains the actions of micro-actors:

There are of course macro-actors and micro-actors, but the difference between them is brought about by power relations and the constructions of networks that will elude analysis if we presume a priori that macro-actors are bigger than or superior to micro-actors. (1981, 280)

Turning his attention to bureaucracy, Latour has argued that material/statistical practices involved with record-keeping are particularly important ways in which macro-actors are assembled out of otherwise disparate human and non-human micro-actors. While his previous work with Callon focused on the construction of a Leviathan-like
actor who could convince a multitude of other figures to act in unison, his later work focused on less material micro-actors. Specifically, he questioned the constitution of abstract “macro-actors that naturally dominate the scene: Corporation, State, Productive Forces, Cultures, Imperialisms, ‘Mentalités’, etc.” Latour’s criticism of these entities was their taken-for-granted nature in determining the actions of micro-actors:

The problem is that these entities could not exist at all without the construction of long networks in which numerous faithful records circulate in both directions, records which are, in turn, summarized and displayed to convince. A “state”, a “corporation”, a “culture”, an “economy” are the result of a punctualization process that obtains a few indicators out of many traces. In order to exist these entities have to be summed up somewhere. (Latour 1986, 28-9)

Michel Callon has continued with this premise in his social studies of financial markets, using the term “economicization” to refer to the way in which the world – including people, materials, practices, and so on – is made knowable in economic terms according to economic theories. As he has bluntly stated, there is “no economy without economics” (Callon and Caliskan, 2)

In his studies of scientific practice, Latour has continued with this methodological premise, developing a concept called “centers of calculation” (Latour 1987, 232) to describe the way in which many modern sciences – from natural sciences to those like economics and demography – generate knowledge. The key aspect of this process is that the disparate, fragmented elements that comprise a macro-actor (a population is his key example) are successively constituted in a form that is increasingly mobile. Through a cascading chain of translations, as he calls them, the surveys collected by a group of census workers are aggregated into a summary table by their local leader, these
summaries are themselves aggregated and abstracted, and eventually the figures are made
into new charts by demographers at the national population office. Once this process is
complete, the director can authoritatively speak about the population in an way that few
others can, precisely because he or she is removed from reality: “demographers can see
things on the final curve summarizing the census (for instance age pyramids) that none of
the pollsters, none of the politicians, and none of the interviewed people could see
before” (Latour 1987, 236). In addition, Latour stresses that the process of translation is
strongest when it is two-way: a demographer can not only create a new statistic from
many reports, but can also refer back to those reports if an objector claims that the
statistic was wrong, miscalculated, or otherwise in error. With enough archival work, the
chain of translations can reach all the way back to the original item from which the first
element was collected, and a claim that has such an archive behind it is far more resilient
than one that does not.

Another example of this process is found in the work of Edwin Hutchins, who
describes the socially distributed cognition with respect to ship navigation. In Cognition
in the Wild (1996), informed by ethnographic research on board a U.S. Navy ship, he tells
of the staggering amount of informational and cognitive work must be performed in order
to keep the ship on course at any give time. In order to cope with these requirements,
information gathering and processing is distributed to crew members, who regularly
collect data, analyze it, and pass the results to others. Hutchins’s research directly
opposes that of cognitive scientists and others who believe that cognition occurs solely in
the minds of individuals. Instead, much cognitive work is distributed, and
“because the cognitive activity is distributed across a social network, many of these internal processes and internal communications are directly observable” (128) – a situation that is not usually the case if one confines cognition to the brain. In this respect, neither the information processing nor the social conditions necessary for such processes “could have been created by the navigator alone” (165). A first glance at something like a Navy ship may give the illusion of natural regularity, but Hutchins repeatedly emphasizes the sociality of such cognitive systems.

Yet Hutchins does not deny the role of the material. In fact, he goes into detail regarding the way in which certain objects, specifically navigation charts, allow their users to perform computationally complex calculations through simple activities. A skilled navigator may be able to keep and alter a ship’s course in his or her head, but nearly any individual who can use a protractor can do so with the right chart. If the ship’s current course will not bring the ship to the destination, all that must be performed to determine the necessary course change is measurement of the angle formed by the current course and a line from the current location to the destination. As he describes, “cognitive abilities that navigation practitioners employ in their use of the forms and inscriptions are very mundane ones – abilities that are found in a thousand other task settings” (131). Furthermore, a chart can be both understood and extended by multiple individuals, which is not the case with the proverbial mental navigator. Because of this, distributed cognition is achieved due to the “general framework onto which specific observations that are local in time and space are projected” (165). Insofar as this framework is sustained by the actions of all participants, information about the ship
is made available to those who need it. However, it must be stressed that distributed cognition is not the same as information sharing, as the crucial contribution Hutchins makes is the role of each human and non-human entity involved in the collective analysis of such data.

In all, Hutchins characterizes social organizations – from the entire ship to a small team – as “computational machine[s]” and claims that “aspects of the behavior of the system can be interpreted in a computational framework” (185). Such systems are powerful and well-functioning even when their members are seemingly in cognitive imbalance. In the various navigation groups in the U.S.S. Palau, he claims that “organized groups … have cognitive properties that differ from those of the individuals who constitute the group” (200). Yet as a well-oiled cognitive machine, the group can perform cognitive tasks with far more speed and complexity than they could otherwise. Hutchins’s main point in his work is that “social organization … does have cognitive consequences that can be described … [through] the observed structures of organizations” (261) as they collectively work out solutions to problems. Furthermore, he stresses that this work must be studied in “the wild” as opposed to in a laboratory. For Hutchins, “culture is a process” (354), and distributed cognition is a primary function of culture, both for modern and non-modern cultures.

The work of Latour, Callon, and Hutchins share a common interest into the material processes through which individuals come to individually and collective know their world. Whether the target of inquiry is a scientific fact, an economic index, or a ship’s trajectory, these authors show that ways of knowing are far more complex
than one may initially believe. In addition, there is a somewhat critical edge to such research, focused on revealing the role of such practices in the maintenance of sociological and technological macro-actors. In a study of vandal fighting in Wikipedia, this account will interrogate the conditions through which editors know both their wiki and their world.

**Vandal Fighting: The Case of 72.68.228.176**

**Technologies of Knowing**

In a wiki as large as Wikipedia, it seems daunting that even a small number of these edits would be reviewed by Wikipedia’s editors, no matter how numerous or committed they may be. However, the default configuration of MediaWiki, the software upon which the project runs, provides a few tools that assist users in such a task. The most visible of these tools is the recent changes page, a dynamically-generated list of all edits made to Wikipedia. This page can be filtered according to a small set of criteria, and must be continually refreshed in order to see the most recent edits. While it is overwhelming to review even a fraction of edits made to Wikipedia, the key functionality of the page is that it lets user review a single edit to a page. This is a departure from the traditional mode of editing in which a work is reviewed as a whole: those patrolling the recent changes page only review what has been changed in a distinct edit. Such a mode of content assessment is made possible by various mechanisms through which an edit is serialized and summarized, reducing it to a few key comparable factors: title of the page, time of edit, type of page (article, talk page, etc), user information, an edit summary (if provided by the editor), and the net number of characters added or removed.
The most powerful mechanism through which this form of editing is made possible is the diff link: a unique and permanent hyperlink that is created for every edit made to Wikipedia. When clicked, the software gathers information from various databases about the edit in question as well as the prior edit, and then dynamically generates a page which allows for quick and easy comparison (Figure 4.3). Users are only presented with sections of the page that were changed, and differences between the two are displayed in red. Among Wikipedians, diffs are the predominant mechanism for representing an edit, even though an edit is, according to the software, only a single revision of a page. As such, diffs combine an edit with its predecessor, transforming a single editorial act into a contextualized record that can be circulated, aggregated, compared, and, as several of the links demonstrate, used to act in various ways. The recent changes page is not the only way to arrive at a diff link.
Another tool that provides a similar functionality is the watchlist. When browsing pages, registered users can click a link to flag certain articles, placing them on a personalized watchlist. Whenever this page is viewed, the software dynamically generates a list of recent changes to only those articles. The watchlist presents edits in the same manner as the recent changes page; the only major difference is the personalization. A third mechanism by which users can arrive at diffs is the
history feature, which is available for every page on the wiki. By clicking the corresponding tab – which is at the top of each page, next to the “edit this page” link – users are presented with a reverse-chronological list of all edits made, including user names, edit summaries, timestamps, and diff links. Such a feature is frequently used if users wish to determine which user made a specific edit to a page. Finally, a similar list can be generated for specific users, displaying all edits made to Wikipedia using a single account or IP address.

**A Vandal Emerges**

At 21:20 on 19 February 2009, a user with the IP address of 72.68.228.176 edited an article on Wikipedia about an unreleased rap album: “Before I Self Destruct” by 50 Cent. Over the next ten minutes, the user made five edits to the track listing and guests section, swapping and replacing a significant number of titles and guests. One of these changes removed the track "Do What It Do" and added "Munch On My Penis" in its place. Next, the anonymous user continued to edit articles about unreleased rap albums, the next of which being “The Mirror (album)” by Ja Rule. At 21:32 and 21:33, two edits were made: the first replacing “Ja Rule has been quiet within music since the album was leaked” with “Ja Rule has been quiet within music since 50 Cent put an end to his career” and the second replacing a sourced blockquote by Ja Rule with the text “I hate 50 cent. He's so mean.” Then, at 21:36, the anonymous user edited the article on “Dr. Dre,” adding “He will bury you if you look at him the wrong way. He wants trouble” to a section about the rapper’s early life.
At 21:37, user DiverseMentality clicked a diff link for the anonymous user’s edits to the article on “The Mirror (album)” and was presented with the diff shown in Figure 4.4. As discussed, there are multiple mechanisms by which users can arrive at such comparative pages, and it is unclear how this user – who declined an interview – arrived at such a link. However, records indicate that he did review 72.68.228.176’s edit to this article using a diff link. On this page, DiverseMentality was presented with two-columns that showed what 72.68.228.176 had inserted and removed in the last edit, allowing him to identify which changes were specifically inserted by this user in this edit. The diff link also allowed him to see that this was the second edit made by this user to this article in a
row, and the “Previous edit” link could have been used to generate a similar diff for the anonymous user’s first edit. While it is unknown if this first edit was reviewed, records show that DiverseMentality clicked the [rollback (VANDAL)] link at the top of the right-hand column for the second diff. This link, which is available to him because he has added an unofficial user-interface modification called Twinkle, resulted in both of 72.68.228.176’s edits being removed, rolling back the article to the version before the user edited the page. After this revert was performed successfully, the Twinkle script navigated DiverseMentality to the anonymous user’s talk page, allowing him to see all previous messages made to this user and send messages of his own. While he could have written a message manually, that is, by typing in a warning about the vandalizing edits, DiverseMentality instead clicked the “warn” tab at the top of the page. This was another feature provided by the Twinkle script, which popped up a box where he could choose from a list of templated warning messages, each with four levels of severity (Figure 4.5). He chose a first level under the general vandalism category, typed the title of the article into the field, and clicked submit. There was space for a customized message, but it was not necessary in this case – the template contained all the information necessary to warn the vandal, shown in Figure 4.6. Of note is the fact that Twinkle, like most vandal-fighting programs, added an invisible comment marker to the message, identifying it as a first level warning. The marker was in the text of the page and could be seen by any user who edited it or viewed its source; however, a viewer would not otherwise be able to know that the text <!-- Template:uw-vandalism1 --> was in the comment.
Figure 3.5: DiverseMentality warning the anonymous user using Twinkle (simulated)

User talk: 72.68.228.176

From Wikipedia, the free encyclopedia

Thank you for experimenting with the page History of the Jews in the United States on Wikipedia. Your text worked, and it has been reverted or removed. Please use the sandbox for any other tests you may want to do. Take a look at the welcome page to learn more about contributing to our encyclopedia. — djhpsode 23:09, 20 November 2006 (UTC)

February 2009

Welcome to Wikipedia. Although everyone is welcome to make constructive contributions to Wikipedia, at least one of your recent edits, such as the one you made to The Mirror (album), did not appear to be constructive and has been reverted. Please use the sandbox for any text edits you would like to make, and read the welcome page to learn more about contributing constructively to this encyclopedia. Thank you. DiverseMentality 21:37, 19 February 2009 (UTC)

If this is a shared IP address, and you didn’t make the edit, consider creating an account for yourself so you can avoid further irrelevant notices.

This is the discussion page for an IP user, identified by the user's numerical IP address. Some IP addresses change periodically, and may be shared by several users. If you are an IP user, you may create an account or log in to avoid future confusion with other IP users. Registering also helps your IP address.
At 21:37, administrator J.delanoy was using a program called Huggle to monitor recent contributions to Wikipedia in real time (Figure 4.7). In Huggle, edits made to Wikipedia appear as diffs and are presented in queues, dynamically ranked according to various pre-defined algorithms. Users can provide their own criteria for queues, but the default queuing mechanism is “Filtered edits,” which displays nearly all edits made to Wikipedia and ranks the “most suspicious” edits first. Additional information is aggregated from various sources inside Wikipedia, such as a user’s talk page or a page’s revision history, in order to contextualize a particular edit. For example, Huggle is able to determine if a user has been previously warned by retrieving his or her user talk page and searching for the previously-described template markers. The program can also detect edit reverts, whether they are made manually, by Huggle, or another vandal fighting tool like Twinkle. All of the information is integrated into the queue: edits with
a yellow box next to their title signify that it was made by a user whose previous edits had been recently reverted; colored boxes indicate users who have been warned on their user talk page using a templated message of the displayed level; brown boxes are for anonymous users; blue boxes are for trusted users; turquoise triangles are edits that revert another edit. Because of the standardized marker, DiverseMentality’s first-level warning was recognized by Huggle, even though it was made using a different program. Because of this, Huggle displayed this information (and more) to all users whose queues showed 72.68.228.176’s edits.

When he advanced his queue to the next edit, J.delanoy was shown the diff of offending edit for the article on “Dr. Dre,” along with 72.68.228.176’s user information. In particular, a green “1” appeared next to the edit, signaling that this user had been recently been given a level one warning for vandalism. By clicking on a red circular button, J.delanoy reverted the article back to a version by a registered user, but did not choose to add a warning like DiverseMentality did. If he did, the red circular button with a triangular sign would have performed both of these functions in sequence. Advancing his queue again, J.delanoy was presented with another one of 72.68.228.176’s edits, this time the last of his five edits to “Before I Self Destruct.” With the click of a button, he reverted all of them, rolling back the article to the version before the anonymous user made the string of edits. J.delanoy is a frequent Huggle user and vandal fighter, making a significant portion of his contributions to Wikipedia in this manner. This was J.delanoy’s twentieth revert that day, and he would go on to make over 180 more during the next four hours. In the prior minute [21:36], he had used this program to revert edits for three
vastly different encyclopedia articles: “Joe Coto,” “Economy of the Republic of Ireland,” and “Billie Holiday.” In two out of these five cases, he chose to add a warning message to the editor’s user talk page; in both of the edits made by 72.68.228.176 and one additional edit, he merely chose to revert.

The anonymous user was not deterred, and made another edit at 21:43 to an article about a rap album, “808s & Heartbreak,” removing an entire section. This edit was placed into the filtered queues of many Huggle users, as the software views mass removal of content by anonymous users whose previous edits have been reverted as highly probable vandalism. However, in J.delanoy’s queue it appeared near the top, because he had previously reverted an edit by this particular user. When the diff (Figure 4.8) appeared in the Huggle window, it clearly indicated that the edit had removed a significant amount of content without adding any. This was quite visible in this form, but it may not have been as blatant if the article was reviewed traditionally, that is, by a user proofreading the entire article after the user removed the section. Less than a minute later, he advanced to the next item in his queue, which was 72.68.228.176’s edit to “808s & Heartbreak.” After reviewing the diff of the edit, J.delanoy clicked a button which instantly reverted it and left a pre-formatted warning message on the anonymous user’s talk page, similar to the one that DiverseMentality made using Twinkle. Before adding a warning message, the Huggle program examined the anonymous user’s talk page and found DiverseMentality’s recent warning, identifying it as a level one comment due to the marker left by Twinkle. In accordance with its programming, Huggle automatically issued a second level warning to the user that was slightly stronger in tone than
The album was recorded in under two weeks inöntew. As implied by its title, “808s & Heartbreak” prominently features the Roland TR-808 drum machine. West felt that the TR-808 was an important instrument that could be used to evolve music, so the concept was introduced to him by Jeff Bhasker. According to West, the fact that Hawaii’s area code was “808” was coincidental, as he had already developed the album's title before seeing the area code in a news article about Vanilla Ice.

The album's lyrics revolve around the voice audio processing technology of the 808. West described the album as being “about love” and “the way love is in the 808 world.”

The album was released on November 23, 2008, by Interscope Records, under the label name “808s & Heartbreak.” It contains 11 tracks, including the popular singles “808s & Heartbreak” and “Eyes on It.”
the previous first level comment. This message also contained a hidden identifier, <!-- Template:uw-huggle2 -->, the significance of which will be later described:

The recent edit you made to the page 808s & Heartbreak constitutes vandalism, and has been reverted. Please do not continue to make unconstructive edits to pages; use the sandbox for testing. Thank you. J.delanoy gabs addt 21:43, 19 February 2009 (UTC)

However, the anonymous user was not dissuaded, making another edit two minutes later to same article, adding the phrase “Kan'Gay west cut his penis off and grew a vagina when he recorded this album” to the end of the “Critical Response” section at 21:45. In a matter of seconds, a bot named ClueBot examined this edit, finding it to be a clear-cut case of vandalism based on its identification algorithms, which include a list of words common to such edits. ClueBot reverted the edit before any other user was able to act, and, like DiverseMentality and J.delanoy, made its way to the anonymous user’s talk page. It received a list of all previous messages left by other users, and after identifying J.delanoy’s message as a second-level warning and DiverseMentality’s message as a first-level warning, issued an even stronger worded third-level warning:

Please stop. If you continue to vandalize Wikipedia, as you did to 808s & Heartbreak, you will be blocked from editing. Your edits have been automatically marked as vandalism and have been automatically reverted. If you believe there has been a mistake and would like to report a false positive, please report it here and then remove this warning from your talk page. If your edit was not vandalism, please feel free to make your edit again after reporting it. The following is the log entry regarding this vandalism: 808s & Heartbreak was changed by 72.68.228.176 (u) (t) making a minor change with obscenities on 2009-02-19T21:45:51+00:00. Thank you. ClueBot (talk) 21:46, 19 February 2009 (UTC)
This task completed, ClueBot moved to examine another edit for blatant signs of vandalism. This was the 592829th edit that ClueBot had reverted since it had begun operation in August of 2007.

Yet the anonymous user was not finished with the page, and added the phrase “KanGay west is the proud owner of a vagina” to the end of the same section at 21:46. A minute later, a user named Eric-Wester advanced his Huggle queue and was presented with this edit, which appeared high on his queue due to the content of the edit as well as the user’s previous warnings. Finding the edit to be vandalism, Wester clicked the same button that J.delanoy used on his Huggle interface to revert the edit and warn the user. The Huggle program, seeing ClueBot’s third-level warning on the user’s talk page, automatically issued a fourth-level warning in Eric-Wester’s name. This message presented the user with an ultimatum:

**This is your last warning.** You will be blocked from editing the next time you vandalize a page, as you did with [this edit](#) to 808s & Heartbreak. Eric Wester (talk · contribs · email) 21:47, 19 February 2009 (UTC)

This final warning was ignored as the anonymous user edited the “808s & Heartbreak” article again, changing the length of the track from “52:05” to “52:05FUck.” This appeared on the top of Eric-Wester’s Huggle queue with a deep red “4” icon indicating that this user had received a fourth-level warning (Figure 9). He clicked the same button to revert this edit and warn the user as he did for the previous edit. When Huggle examined the user’s talk page, it found that Wester’s previous message was a level four warning – the highest. The program asked him if he wanted to report the user to Wikipedia’s administrators, which would take the form of a request to
temporarily block this IP address from editing. He clicked “Yes” and the Huggle software made an edit to a different page: “Wikipedia: Administrator intervention against vandalism” or AIV. After collecting the edits which had been marked as vandalism in the level three and four warnings (issued by him and ClueBot), Huggle generated a templated report. Similar to the warning messages, the report was generated in a standardized manner: all a human or non-human editor has to do is provide a few variables, and the necessary text would appear to all viewers. Huggle edited the AIV page and appended the report below four others, some of which had already been reviewed (Figure 4.9).

Administrator Intervention against Vandalism (AIV)

The most visible aspect of the reports is the numerous links included in all of them, which is a result of the template used to generate standardized reports. When the
text {{IPvandal|72.68.228.176}} is inserted into a page on Wikipedia, the twelve links that begin a report are automatically generated. Any user or administrator who visits AIV can click any of these links to have certain software programs gather data from various sources and generate a report about this anonymous user. The “talk” link is to the user’s talk page, where all the warnings were made; “contribs” provides a list of all the edits made by this user; and “deleted contribs” gives administrators a list of all edits made which were deleted. The next five links are to various external, non-Wikipedia

![Figure 3.10: AIV page, at the time of Eric-Wester’s report (truncated)](image-url)
tools that aggregate publicly-available information about IP addresses, including approximate geographical location, Internet Service Provider, and more. The “block user” link takes an administrator to a form which allows him or her to issue a technical block on editing, while the “block log” link gives a list of previous blocks made to this user. The rationale follows the links, which for Wester’s report included three diff links to incidences of vandalism, selected and inserted by Huggle. Wester’s signature concludes the report, with links to his talk page, contributions, and e-mail, as well as a timestamp.

Before this report was reviewed by an administrator, the anonymous user made another edit, this time to the article “Conan O’Brian.” He replaced the occupations of his parents with “pizza delivery man” and “majure in flamingos” and appended “If you look at him the wrong way, he will kill you in a jiff” to the end of the introduction. Appearing high on Wester’s Huggle queue, Eric Wester reverted this edit the minute it was made, clicking the ‘revert and warn’ button in the same way that he had done with the previous edits. Instead of warning the user or issuing a new report, the program identified that an existing report had been made on AIV and amended it, adding a diff link of this edit to the list of vandalism incidences. Because of a bug in the software, the links appeared as “1 2 3 3,” but the diff link to the new incidence of vandalism was still accessible from this mis-numbed addition.

The anonymous user continued to edit, making the same changes to the “Conan O’Brian” article that were just reverted by Eric-Wester. In a matter of seconds, this edit rose to the top of J.delanoy’s queue for obvious reasons: the edit was previously
reverted seconds before; it was made by a user who had not only been reverted multiple
times, but reported at AIV; and it included both a misspelled word (“majure”) and a word
closely associated with vandalism (“kill”). The most blatant of these was 72’s status as a
reported vandal, and in Huggle queues, the icon next to the edit appeared as a red circle,
indicating that it was made by a user who had been reported at AIV. When J.delanoy
clicked the ‘revert and warn’ button on Huggle, the article was rolled back to the version
last edited by Eric-Wester, like it would have been for any other edit identified as
vandalism in this manner. Normally, the Huggle software would have asked J.delanoy if
he wished to block the user, given that J.delanoy was an administrator and the
anonymous user was reported at AIV. However, Huggle’s blocking code had been
recently disabled because it was undergoing maintenance, so J.delanoy had to block the
user in another way.

In a web browser with the Twinkle script installed, J.delanoy navigated to the
anonymous user’s talk page and clicked a “block” tab, which was available to him
because he had the Twinkle script installed. A box popped up and asked him to choose a
block length (48 hours) and a rationale (“Vandalism”), then confirmed the action.
Twinkle then sent a command to the Wikipedia software, authenticating as J.delanoy and
issuing a request to insert a single row to a database called “ipblocks.” After confirming
the technical validity of the request, 72.68.228.176’s temporary block was inserted into
this database, along with information such as J.delanoy’s user id, his rationale, the current
time, and the time when the block would expire. Next, Twinkle navigated to the user’s
talk page and asked J.delanoy to select a templated block message, similar to
DiverseMentality’s warning. This message was told the user that he or she had been temporarily blocked from editing Wikipedia for “abuse of editing privileges” (Figure 4.11).

Under a minute later, a bot named HBC AIV Helperbot3 requested a list of
recently banned users and the current content of the AIV page. The MediaWiki software then queried the recently-updated ipblocks database and retrieved 72.68.228.176’s block – among others – and passed both this list and the content of the AIV page to the bot. The bot parsed through all the vandalism reports and matched the IP address identified in Eric-Wester’s report with the recent block by J.delanoy, which indicated that the request to block the user had been fulfilled. As it had done tens of thousands of times before, the bot removed the report from the AIV page, as it was now unnecessary for the purposes of vandal fighting. Yet because of the history feature built into every wiki page, the report was not permanently removed from AIV, as any user could go back to the version of the page before the bot removed the request – as has been performed in order to retrieve the data presented in this chapter. Two hours after the anonymous user’s 48 hour block expired, another edit was made from the same IP address to the article “Don Cannon,” appending: “Don Cannon is known to wipe his butt to quickly, which results in some small pieces getting stuck in his ass.” Erik9 reverted this edit, navigated to the user’s talk page, saw that the user had been previously banned, and clicked the “report” button, which was provided to him because of a script similar to Twinkle called ARV. He typed in a rationale (“more vandalism soon after last block”) and clicked “submit,” which was transformed by ARV into a well-formed block request that was subsequently posted to the AIV page. Eight minutes later, administrator PhilKnight reviewed Erik9’s report and, using a “block” link provided to him using Twinkle, blocked the anonymous user for an additional 48 hours and left a templated message on the anonymous user’s talk page informing him or her of the block, similar to J.delanoy’s.
Analysis

The most immediate significance of this account of vandal fighting is the extent to which technological tools are shown to streamline the reviewing of edits and the identification of malevolent users. Like Hutchins’s analysis of navigational charts, technological actors like Huggle make what would be a cognitively-intensive process into a mundane affair that can be performed by users who may have little to no knowledge of the specific topic at hand. For example, a significant portion of vandalism made by the anonymous user is clearly identifiable as such, even if one knows nothing of hip-hop. The most obvious cases of this are insertions of obscenities/nonsense and mass removal of content, which are almost always vandalism. However, such obviousness is not an a priori condition, but rather an achieved state: the edits in question were made more suspicious because they were presented in this chapter using the same kinds of tools and programs that vandal fighters use in their daily work. While nonsense or obscenity may be easy to spot when reviewing the page traditionally – that is, proofreading the entire article – removal of entire sections is a common form of vandalism that is difficult to detect by merely reading the article. However, through various tools which abstract and re-present edits, both the insertion of obscenities and mass removal of content are equally visible, albeit in various ways. The diff is one way in which an edit is contextualized and visualized, allowing a quick and easy comparison between only that which is changed. – whether the edit removed entire sections or inserted a single, word that changed the meaning of the sentence.

The Huggle program is another way in which edits are further
transformed, contextualized, and abstracted. The built-in queuing mechanism, which by default ranks edits according to various vandalism-identification algorithms, is a form of delegated cognition. Users of Huggle’s automatic ranking mechanisms do not decide for themselves which edit they will view next. Instead, the software gathers as much information as it can about each edit in the queue and then gives the user the most likely candidate for vandalism. Users who use alternative queues, such as the misspelled words queue, are involved in a similar relationship with the software, one that does not strip the user of his or her agency. Instead, the vandal fighter reserves his or her cognitive duties for those which cannot be replicated by computerized algorithms. Because of the complexities involved with misspelled words and vandal reports, the final determination to act remains made by humans – in theory, at least. As may be expected, controversies have emerged over both spell-correcting bots and so-called adminbots, which performed these tasks without much discretion and to much criticism.

Yet the most notable aspect of the Huggle software – as well as Twinkle, ARV, and other assisted editing tools – is the way in which they collectively enable a form of distributed cognition among otherwise disconnected vandal fighters. As was made apparent, a significant number of human actors are required to act in coordination with each other to ban a vandal. In this case, four human editors each made separate determinations of vandalism within a fifteen minute period, which ordinarily would not be sufficient to make such a determination. However, through the specific software used by the editors, identified incidences of vandalism were reported to the user’s talk page, which was used more as notice board or data base for other vandal fighters than a
space for the anonymous editor to discuss issues. This shows that the process of warning is not to merely inform a user that their actions are disrespectful or unwanted, but rather one of translation. Once a user has reviewed an edit and determined that it is vandalism – using any number of mechanisms or tools – this cognitive work is preserved in the form of a warning. The edit, which had been previously made circulatable in being transformed into a diff, is further abstracted and contextualized by incorporating it into a warning template. Even if a user performs this task manually, that is, by appending the text

```
{{uw-huggle1|http://en.wikipedia.org/link-to-diff|Article Name}}
```

to a user’s talk page to create a first-level warning, the chain of translations will be furthered. Through the user talk page warning, a record of vandalism will be created that can be subsequently deployed by any vandal fighter, administrator, or other interested user. If previous edits were identified as vandalism in this manner, other users do not have to trawl through the user’s recent contributions: unassisted vandal fighters can visit the user talk page to see previous warnings, and assisted users can simply have the software automatically incorporate this information into its decision-making process.

The use of hidden comment codes for marking templates with a severity level is essential in this ability, as it compresses all the contextual information associated with an identified incident of vandalism into a single number, one through four. Even though hundreds of warning templates exist (and more are added regularly as vandal fighters disagree on their wording or formatting), the standardization into levels of severity allows a warning to be operationalized, circulated, aggregated, and incorporated into a process of distributed cognition. With a program like Huggle, which can determine if a user
has been sufficiently warned by the vandal fighting community and report those who continue to abuse their editorial privileges (a rather subjective determination), the key feature of a Latourian translation chain is revealed: reversibility. In generating an AIV report, Huggle navigates back down the translation chain in order to gather the diff links which had been previously inserted into warnings.

Bots also make this process of distributed cognition possible. The key bot in this story appears to be ClueBot, who reverted one of the anonymous user’s edits and contributed to the process of vandal identification. However, there were an additional set of bots performing key infrastructural work, known as the AIV Helperbots. These automated editors maintain the AIV queue in an orderly fashion, removing reports that have been filled, append relevant information to certain reports, and let interested users know when there is a current backlog of requests that needs to be filled. These bots are critical organizational actors in the vandal fighting process, and it is quite likely that AIV would fall into disarray without such invisible maintenance work. Together, the AIV Helperbots edit AIV once every five minutes, on average, although this figure obviously varies with the level of reports issued and responded to. The strongest evidence for the infrastructural status of these bots is their multiplicity: no other bot has multiple instances of itself, while there are currently four distinct instances of the AIV Helperbot – all running the same code – currently active. These bots are also run from geographically distributed locations (New York, Germany, Australia, and the Netherlands) to “reduce the chance of there being no bot to perform the task if one were to fail,” as its operator has claimed.
AIV Statistics

The case of 72.68.228.176 is an exemplar of vandal fighting. As Figure 4.12 shows, many users and IP addresses are routinely blocked each day, ranging from five hundred to twelve hundred on any given day. As the data point for 4 February 2009 makes clear, there are a few days in which a relatively high number of users are blocked.

Figure 3.12: Blocks per day
However, this particular incident was due to the identification of a large number of open proxies; these ‘anonymizer’ services allow any user to mask their IP address. Because of the potential for abuse, all known open proxies blocked from anonymous editing as a matter of policy. Furthermore, the activity of warning using automated and assisted editing tools is significant, as Figure 4.13 shows. While warnings peak during the middle of the week and trough during weekends, the observed level of warnings has not dropped below three thousand per day. As expected, the frequency of each warning is inversely-proportional to its severity.

Figure 3.13: Warnings per day by automated or assisted editing tools, stacked by level
Yet despite the regularity of the vandal fighting process, it would have been nearly unrecognizable in 2005, when the AIV page was first used to coordinate blocking requests. The growing deployment of bots and assisted editing tools can be seen in statistics generated from a list of all edits to AIV, which were retrieved using the same bot program described in the beginning of this chapter. While bots can be easily identified using their usernames, detecting the work of programs like Huggle and Twinkle are more difficult. By default, most assisted editing tools add a unique marker to the edit summary, such as “[[WP:HG]]” for Huggle and “[WP:TW]” for Twinkle, simplifying automated coding of edits. Some tools – most notably Huggle – did not always incorporate this feature, but most left complex comments and edit summaries in a programmatic manner. Such indicators are easy to detect via automated coding, but difficult for a human to reproduce; these edits were coded as assisted edits made by the program “unknown.”

The first graph in this set (Figure 4.14) shows the proportion of edits to AIV by assisted edit status. Bot edits are included, although the only noticeable bots are the HBC AIV Helperbots: their introduction into AIV in early 2007 is clearly visible, and they have made about half of all edits since that time. The declining popularity of early vandal fighting tools, such as ARV, is also visible, as is the corresponding increase in popularity for Twinkle and later Huggle. The light blue area represents the previously-discussed “unknown” tools, suspected to be Huggle. The stark change in summer 2008 from unknowns to Huggle was due to the program inserting the unique [[WP:HG]] marker into edit summaries by default – previously, such a feature had to be
explicitly enabled. In all, this graph shows that the popularity of assisted editing tools
and bots has grown dramatically, and that only about twenty percent of all current edits to
AIV are unassisted.
Figure 3.14: Relative frequency of edits to AIV, stacked by assisted editing program used
The second graph in this set (Figure 43.15) is a panel of all edits to AIV between April 2007 and March 2009, with each data point representing three days. The graphs are stacked by assisted/manual status and paneled by both user type (bot, admin, and user) and edit type (addition, amendment, and removal). This second variable takes advantage of new software functionality added to MediaWiki in April 2007, which began recording the length of the page in edit histories. From this data, the net increase or decrease in page size was calculated by comparing each edit to its predecessor, and edits were classified accordingly. Edits of a negative length indicate removal of content, edits adding from 0 to 100 characters indicate amending a previous report (including extending reports, as Eric-Wester did using Huggle), and edits over 100 characters indicate a new report or comment. Because of the paneling, there are actually nine charts in Figure 13: the top left chart is of administrative additions, the middle left chart is of editor additions, the bottom left chart is of bot additions, etc.

This set of graphs makes visible the different types of edits made by editors and administrators, contesting the view that administrators perform a significant amount of the maintenance and governance work required to keep Wikipedia running smoothly. Far more additions are made to the page by editors than administrators, and assisted editing programs are predominantly used by editors, not administrators. As described, bots predominantly remove content, and interestingly, the level of bot removals and editor additions is roughly equal across time. Another notable observation is that the addition of the AIV report extension feature to Huggle is visible in the user/amend panel. The feature was released in February 2008, which is the first time that assisted amendments appeared in the editor panel.
Figure 3.15: Edits to AIV over 3 day periods, paneled by user type and edit type, stacked by assisted/manual edits.
Conclusion

As this case shows, technological tools like bots and assisted editing programs are significant social actors in Wikipedia, making possible a form of distributed cognition regarding epistemological standards – independent of what those standards happen to be. The network of associations constituted around vandal fighters, administrators, bots, assisted editing tools, diff links, warning templates, user talk pages, and the AIV queue is one through which Wikipedians come to know various facts about their wiki. Such knowledge may not be encyclopedic (or even knowledge at all, depending on various definitions of the term), but are critical to the process of knowledge production within the project. Through such a process, the macro-actor of governance comes on the scene, capable of acting in ways that seemingly-dominate users in order to maintain order in the wiki spaces. However, this account has not taken for granted the stability or unity of such a macro-actor, instead showing the way in which something like governance is made to act through a diverse assemblage of human and non-human actors.

As such, this chapter definitively rejects the ‘wisdom of crowds’ account of Wikipedia’s order, which holds that multitude of unmediated, uncoordinated, independent rational actors chooses the ideal form of knowledge in the same way that a free market chooses the ideal price of a good. While the account of vandal fighting is a form of this ‘collective intelligence,’ such ways of knowing do not spring forth fully formed once a systems administrator installs wiki software onto a server. Instead, the
sociotechnical systems which seem to structure user behavior must be constantly maintained alongside a heterogeneous set of norms, roles, discourses, standards, programs, visualization tools, and databases. There is a staggering amount of work that must be performed in order to maintain and align these overarching structures such that they can have the effects that sociologically-themed research suggests. If, as in many corporate or academic wikis, a single team was tasked with maintaining such a normative-epistemological infrastructure, such administrators would be overwhelmed on the first day. As the statistical data indicated, contributions to AIV are numerous and come from a diverse population – contesting various cabalistic interpretations of Wikipedia’s order, it was shown that a vast majority of vandalism reports are added by non-administrative editors. Yet size of the administrative corps or editorial base is not the only factor in Wikipedia’s success. As this chapter has shown, the cognitive and organizational work of such actors is distributed to both human and non-human actors, who rely on each other to maintain this network of knowing.
Chapter 4. Conclusion

Wikipedia’s open-source production model is the epitome of the so-called Web 2.0, an egalitarian environment where the web of social software enmeshes users in both their real and virtual-reality workplaces.[...] Wikipedia has become a model of what the collaborative Internet community can and cannot do.

- “Wikipedia,” Encyclopædia Britannica

Wikipedia, the self-proclaimed “free encyclopedia that anyone can edit,” is often mystified in the public imagination. One of the most common views is that hundreds of thousands of volunteers, each with their own subjective viewpoints, continually fight to define reality in a relativist battleground. Representations of Wikipedia in academia and contemporary media have reinforced this view, depicting the project as one built – socially, culturally, and technically – to resist organization and management in favor of a political model called “mob rule” by critics and “radical democracy” by supporters. Many critics and supporters alike continue to allege that Wikipedia’s mission is nothing less than an full-on assault on traditional institutions of knowledge production, taking its open model of encyclopedia-building to be a radical and often politicized rejection of expertise in favor a more intentionally-anarchistic conception of editorial authority. In Wikipedia – so the cultural narrative goes – there is a purposeful absence of all the safeguards which have been traditionally deployed to govern the production of knowledge in every sphere from journalism to science.
Given this view of Wikipedia, the project’s unexpected success is one that is difficult for most to explain. As formal and informal evaluative studies continue to show, encyclopedia articles written in this supposedly anarchistic manner are surprisingly accurate. While Wikipedia’s relative quality (compared to more traditional encyclopedias) is a hotly-contested issue, popular and academic accounts continue to express shock regarding not only the project’s size and scope, but the content and coherence of particular articles as well. How could such a radically egalitarian, anarchistic, relativist, and populist model of knowledge production possibly result in anything that even resembled an encyclopedia? Such a question is becoming increasingly relevant as a growing number of journalists, activists, businesspersons, and academics are attempting to replicate Wikipedia’s success to their activities, or at the very least thinking seriously about how the new form of knowledge production will affect their organizations. In almost every enterprise that is even tangentially related to the production or dissemination of knowledge or information, manifestos have emerged declaring the so-called ‘wiki way’ to be the newest paradigm to which we all must shift or quickly become irrelevant. Not merely limited to replicating the literary form of the encyclopedia and related reference works (e.g. dictionaries, directories, atlases, etc), wikis are being heralded as the new platform for journalism, classroom education, scientific research, creative writing, religious practice, music and film production, political activism, and even the legislative process itself. If the wiki model – which is often reduced to the ‘anyone can edit’ functionality of the software system – can be
applied to these cultural activities, it is claimed that the result will be far more productive, efficient, popular, participatory, and egalitarian and far less coercive, biased, hierarchical, elitist, and stagnant.

Yet despite the fact that countless numbers of wiki-based projects have been founded, none have replicated Wikipedia’s explosive size, scope, quality, and popularity and few have achieved a more tempered metric of success. Like the open source software movement to which it is often compared, Wikipedia is frequently characterized as a project that by all accounts should not be working at all, much experiencing continuous growth in terms of both content production and viewership. Harvard Business School professor Karim Lakhania expressed this sentiment rather succinctly when he claimed that Wikipedia “works in practice, but not in theory.” While Wikipedia seems to be immune from vandals, spammers, and ideologues, it is an open question if the project’s success is merely a fluke that cannot be applied beyond general purpose encyclopedias. This uncertainty is all the most prevalent because common sense, sociological theory, and the history of politics all impart a significant amount of skepticism towards the collective abilities of the body politic. As such, a growing number of popular and academic accounts have come on the scene to theorize this new, collaborative model of knowledge production and information sharing. One of the easiest but most problematic ways of dealing with the anomaly of Wikipedia has been to lump the encyclopedia project into the heterogeneous assemblage of Internet-based cultural artifacts commonly referred to as “Web 2.0” – blogs, social networking sites,
media sharing platforms, news aggregation sites, and more – and claim that their collective success is not only self-evident, but emblematic of life in our postmodern, globalized, information age. Under this model, traditional sociological and political theories about how individuals interact with each other in groups are relics that no longer apply; we must simply accept that sites like Wikipedia work without needing leadership, authority, control, centralization, organization, or hierarchies to govern its members.

However, this is rather unsatisfying theoretical explanation, and many have refused to take Wikipedia’s success at face value. In response, a myriad of theoretical explanations have come on the scene in an attempt to claim Wikipedia as an exemplar for their view of how individuals interact in groups to produce knowledge or information. Instead of representing Wikipedia as an entirely new phenomenon in contemporary society that cannot be explained, these researchers have argued that the project is far more familiar to us when it is more closely examined. While nearly all academic disciplines or theoretical traditions within the social sciences have been deployed in such a manner, the most predominant positions are inspired by either traditional sociological theory or social network theory. The more sociological researchers claims that Wikipedia is more like a traditional society, with social rules, roles, orders, norms, institutions, and other structural elements to keep editors in line; in contrast, the social network theorists tend to claim that Wikipedia operates like an economy, coordinating the uncoordinated actions of editors as if they were buyers and sellers in a marketplace.

The sociologically-themed research – which comes from scholars in a number of
disciplines outside of Sociology, including Communication Studies, Computer Science, and Information Studies – tends to focus on elements and aspects of Wikipedia’s social structure and cultural milieu. Such research operates on the assumption that in Wikipedia – as in any society – these sociocultural structures are what prevent Wikipedia from degenerating into anarchy, despite its open editorial model. One of the key reasons these researchers give for Wikipedia’s unique success among all other wiki-based projects and previous forms of encyclopedia-building is this common lifeworld in which participants are embedded from the moment they enter as an editor. Research in this tradition generally describes the way in which participants in the project are guided by a diverse assemblage of norms, rules, roles, classes, workflows, institutions, governance mechanisms, power relations, or discursive orders. A significant amount of research in this tradition has quantitatively revealed various demographic and sociological qualities about the project’s editorial base, including statistics regarding the distribution of labor, inequality of contribution levels, and enforcement of codified norms. In contrast to popular depictions, the sociologically-themed research often claims that the project is not that different from many off-line societies or social groups, guided by strong norms and social structures that are invisible to the casual editor or observer. Yet once this “hidden order of Wikipedia” is revealed, the unexpected success of the global society of editors appears far more predictable.

In contrast, research from social network theorists tends to reject the view that strong sociocultural elements keep Wikipedians in line; instead, they analyze the
encyclopedia project in the same way that economics analyzes a marketplace. By reconceptualizing editing in Wikipedia as a ‘transaction’ between two individuals (the individual making the edit and the individual whose edit is modified or removed), such research has crafted powerful explanations of Wikipedia’s emergent order. In the same way that the ideal price of a good emerges out of the micro-level interactions between a multitude of buyers in a well-functioning market, so does the ideal version of an encyclopedia article emerge in Wikipedia. While any given transaction or edit may be ‘irrational’ (unduly influenced by ideology, coercion, ignorance, or emotion), the networked technology that allows any individual to contest the actions of any other individual ensures that articles generally remain encyclopedic. However, as network theorists emphasize, this effect is directly related to the size and constitution of the project’s editorial base: there must be enough editors for the process to achieve critical mass, and the majority of editors must share at least a nominal vision of encyclopedia-building. If a predominant number of contributors view the wiki as a graffiti wall or advertizing board – as in my wiki I described in the introduction – a wiki administrator should expect nothing less than the ‘wisdom’ of these contributors to triumph in the wiki model of editing. Yet according to these theorists, because Wikipedia contains enough dedicated individuals who share a weakly-normative abstract view of how their project ought to operate, an emergent order arises not from a top-down social structure enforced by elites, but instead out of the hundreds of thousands of micro-level negotiations which occur every day. Other similarly-minded theorists portray this libertarian view of
Wikipedia with biological metaphors instead of economic ones, invoking a Darwinian ideal of natural selection to explain how good contributions thrive and why bad contributions quickly die out. However, they remain united in a libertarian ideal in which the rationality of Wikipedia is found not in criteria and mechanisms arbitrarily designed and imposed by a global authority, but rather in a naturally occurring process that operates largely at a local level.

Both the sociological and network theory perspectives focus on a different aspect of Wikipedia, and their near-exclusive focus often leads them to privilege certain elements and ignore others. Sociologically-themed research into Wikipedia tends to overstate the role of overarching social orders and cultural milieus in structuring or determining human behavior, taking for granted the stability and unity of these superstructural elements. Similarly, social network theory tends to problematically attribute the emergence of order to the phenomenon of networked interaction, often ignoring the strong roles that Wikipedia’s governing institutions and actors play in the maintenance of social order. Yet as I have shown in this work, the issue at hand is not merely the fact that neither macro-level social phenomena nor micro-level emergent properties of networks fully explain Wikipedia’s unexpected order and stability. Rather, both traditions have largely ignored the role of technology in Wikipedia beyond the “anyone can edit” functionality. Specifically, I argue that software agents – many of which are not directly incorporated into the project’s software architecture – have been inappropriately discounted in both social and economic depictions of the project. Both
traditions take for granted the work performed by these technological infrastructures, attributing the phenomenon of social integration to a non-technical macro-actor: emergent properties of well-constituted networks in the case of social network theory and sociocultural superstructures in the case of sociologically-themed research.

My core argument in this work is that Wikipedia had to be made into the kind of place in which either sociologically-heavy concepts of normativity and governance or strictly-atomistic theories of networked interaction could come on the scene to explain human behavior. It must be noted that I am in no way denying phenomenon of emergence in Wikipedia as an explanatory concept – whether it is from a tradition that dates back to Adam Smith’s uncoordinated invisible hand or Emile Durkheim’s objectively real social facts. Coordinate network effects and norm-based self-governance may very well be excellent heuristics for describing the way in which users in Wikipedia interact with each other to collectively author an encyclopedia. While I do favor the more sociologically-themed view over that presented by most social network theorists, this thesis does not primarily intervene on behalf of one side against the other. What I instead contest is the organic spontaneity that researchers from both traditions tend to ascribe to the way in which order is produced out of disorder – regardless of the specific manner in which they theorize the dynamics of editorial action. In my analysis of human and non-human actors working within Wikipedia, I have shown that there is a significant amount of transformations, aggregations, contextualizations, operationalizations, and other processes performed behind the scenes, making possible the kind of interactions
that are most immediately visible when examining the encyclopedia project. In a phrase, emergence does not simply emerge; it can only materialize in the manner described by theorists from both traditions because a vast set of infrastructural elements has and continues to be constantly assembled in rather specific ways. Obviously, giving anyone the ability to edit any page at any time in any manner is a critical element of this process, but it is a mistake to attribute the rationality of networked economics or the unity of sociological structures solely on the basis of this software feature.

It may very well be the case that editors are motivated by incentive structures to produce coordinate effects without needing explicit coordination, or structured by norms to make edits in socially-amenable ways without needing hierarchical governance. What I argue is that there exists a highly-material infrastructure of knowing through which editors are made into the kinds of individuals whose contributions, edits, and other actions can be seen as operating in such a manner largely by themselves. The production and maintenance of order is therefore more mechanistic than sociologically-themed research assumes, but substantially more structured than social network theorists claim. While there is a large amount of social discussion and negotiation as well as non-discursive interaction between disassociated human editors (facts I am in no way denying), Wikipedia is held together by far more than discourse and norms on one side and the emergent properties of networked economics on the other. Reconciling the contradictory views of Wikipedia espoused by these two traditions is therefore not merely an activity of incorporating the activities that the other ignores, as both traditions depict
the activity of editing in a way that fails to take into account the material conditions of possibility for such acts to occur.

In both my analysis of bots and vandal fighting, I have repeatedly shown that while explanatory theories espoused by various researchers may indeed be true, this is only the case because of an active and largely-unofficial technological infrastructure that remakes Wikipedia into the kind of entity which can appear to both editors and observers as operating in line with these assumptions. Far more than technological reifications of what could be called political issues (e.g. behavioral standards, administrative duties) these infrastructures of knowing make Wikipedia a place in which the collective production of encyclopedic knowledge can occur without the process degenerating into a primal anarchism. For example, Wikipedia can appear as a networked economy in which the actions of uncoordinated editors ‘simply’ result in emergent coordinate effects only because there are a significant number of software agents working diligently to produce such effects through the continual transformation, aggregation, operationalization, and comparison of editorial actions. Not knowing the diverse assemblage of actors – human and non-human – involved in the identification and banning of a vandal may lead a network theorist to claim such an incident as a perfect example of uncoordinated emergence. Yet having showed the way in which these software agents must actively work to coordinate uncoordinated users, the rationality of Wikipedia must be ascribed to their actions, not any property of networked economies or crowds.

The same criticism exists of sociological explanations, which relies on social
macro-actors in the same way that network theory relies on properties of networks. Wikipedia can only appear to be a society in which micro-level disputes are resolved by ‘simply’ linking up to macro-level actors like norms, governance mechanisms, and social roles because of all the work that is done by software agents in holding together this shared vision of how Wikipedia is and ought to be. Echoing the ironic cry of AntiVandalBot’s plea for political suffrage presented in the introduction, these software agents are editors too: any attempt to tackle the issue of Wikipedia’s success must take into account these non-human actors as much as they do humans. Yet I am not merely claiming that both sides need to take into account more of the project’s often-unofficial software agents, because these non-human actors are far more than active participants in Wikipedian society. Such actors constitute the conditions of possibility upon which both norm-governed social agents who establish micro-to-macro associations as well as atomistic, uncoordinated individuals who collectively act to produce coordinate effects come to be constituted in Wikipedia. The irreconcilable contradiction between these two views of Wikipedia is therefore based on the faulty assumption that such a view is binary; that is, because of the wiki model or even human nature individuals in Wikipedia are constituted in one way or another.

One group of these forgotten actors are bots – fully-automated computer programs that correct, remove, flag, or otherwise transform edits and articles according to programmed algorithms. Sociologically-themed research that does focus on these non-human editors predominantly characterizes them as force multipliers that are not that
different when compared to social artifacts, like standards or accounts of best practices. Furthermore, in descriptions of dispute resolution or governance acts – in which the local negotiation of content occurs largely based on global norms and practices – there has been a need to posit an already-existing ‘policy environment’ or other social superstructure to guide this enforcement. In contrast, I showed that these software programs regularly and consistently perform a significant amount of social work in holding together normative visions of how Wikipedia is and ought to be. While it may be presumed that there is an invisible social force that guides users to, for example, sign their comments, not vandalize articles, not submit malformed requests to administrators, and so on, a significant amount of this work necessary to socially integrate Wikipedian society is performed by bots. In other words, the observed fact that all comments in a discussion contain signatures should not be taken as evidence of a strong social norm in favor of the practice without taking into account the fact that a bot is unceasingly signing unsigned comments seconds after they are made. More importantly, I have shown that these technological actors make possible forms of social interaction and negotiation about controversial issues beyond the work they perform in maintaining an ideal discursive space against malevolent users. As the case of HagermanBot and opt-out lists showed, a new software feature reconciled two competing worldviews regarding the auto-signature bot, which became the most valuable ally in defending the bot’s actions from detractors. As the case of BetacommandBot showed, this is not a technologically determinist process in which software manipulates human behavior, as the feature of opt-out lists became a
social artifact onto which an editorial right was projected and articulated in a time of controversy. In such debates, these technologies became important participants in negotiating what it meant to be an editor and a bot operator in Wikipedia.

Another class of technological actors systematically ignored by both traditions of Wikipedia scholarship includes assisted editing programs and visualization features – software tools that transform the way in which editors experience their project and those who edit it, making possible a specific way of knowing essential for enforcement of normative and epistemological standards. My analysis of vandal fighting directly contests depictions of the process from both sociologically-themed research and social network theory. I showed that the process of ‘collective intelligence’ – through which a crowd comes to make a wise decision even though its members are seemingly atomistic, independent actors – is only possible because of a strongly aligned infrastructure of knowing. Contrary to the depictions of Wikipedia by social network theorists, neither vandal fighters nor the vandals they fight are disconnected, independent agents; instead, users are always-already associated in largely technological ways by virtue of many unofficial software programs. Software agents in Wikipedia constantly structure the site and actors within it to such an extent that mere human coordination on a user-to-user level is all that is – and needs to be – recognized by observers and casual users. Likewise, the normative structures that sociologically-themed research posits to explain the coordinated interactions among editors may very well exist, but they must not be taken as the primary cause for explaining why certain kinds of edits get reversed. The
sociological structure very much exists as these researchers describe, but it is a product of the endless interactions between vast assemblages of humans, databases, and software agents. Furthermore, the story is not a simple determinism in which technology silently reifies visions of how Wikipedia is and ought to be, but rather one in which these infrastructures of knowing and knowledge production are just as much participants in the activity of performing society as they are pre-forming it. In contrast to Langdon Winner’s famous declaration that artifacts have politics, I have shown that these non-human actors are active political participants in Wikipedia. While they may not be able to run for Wikipedia’s highest offices, these software agents and other related tools are integral members of the Wikipedian society that deserve far more attention than they have received in both popular and academic discourse regarding the encyclopedia project.

Returning to the claim that Wikipedia “works in practice, but not in theory,” it is clear that the inability of scholars to provide an adequate framework for understanding the project is due to the inability to recognize the importance of such non-human actors. Most research takes for granted the stability and unity of the software platform upon which Wikipedia runs (which is largely reduced to the “anyone can edit” functionality), and takes such a feature to be representative of ‘the wiki way’ in terms of technology. Explanations for Wikipedia’s success – which are almost exclusively focused on the way in which individual human users edit and interact – are built on top of this conception of how Wikipedia works. In other words, most Wikipedia research begins with an incomplete view of the process by which edits are made and then attempts to construct
various social scientific theories which would explain why coordination results. The reason why Wikipedia does not work ‘in theory’ is that no plausible theory could be constructed which could explain all the work performed by non-human actors while simultaneously denying their existence. It is an attempt akin to explaining traffic flows in a major city without taking into account the work traffic lights perform. And to expand on the analogy, those who research Wikipedia should be more involved in incorporating these traffic lights in explanations, as opposed to crafting theories that rely on broad cultural expectations of respect or the ad-hoc negotiations between drivers to explain how a multitude of drivers with often-contradictory intentions are able to pass through a busy intersection.

There is a significant amount of further research that needs to be performed along these lines. First, this research only encompasses the English-language version of Wikipedia, and while the English version is the largest and most active, there are many other active language versions. At present, it is unknown the extent to which these same kinds of software agents operate in other version of Wikipedia, and further research is needed on both qualitative and quantitative fronts. Next, I have only given a few accounts of events within Wikipedia in order to describe the active roles software agents play in producing order from disorder in the encyclopedia project. The academic study of Wikipedia could benefit from a series of research projects investigating the roles of other technological actors. The process through which pages are protected or deleted by administrators follows a similar procedure as the banning of vandals, as does the
reporting of users who may be disruptive but do show indications of good faith. These activities ought to be studied to see if bots, visualization tools, and assisted editing tools operate similarly.

While this study has largely focused on unofficial software features, there are many important software features which have been officially incorporated into Wikipedia, such as the previously-mentioned spam filter. The negotiation of such features is a relevant issue for this topic, and one event in particular is especially noteworthy: Flagged Revisions. This feature would have required that some or all edits made to Wikipedia by certain kinds of untrusted users be put into a queue and would only appear to the rest of the world if certain kinds of trusted users approved it. While the German-language Wikipedia incorporated it after a project-wide vote, the dispute over Flagged Revisions has been brewing for almost a year on the English-language Wikipedia and shows no signs of a compromise. The recent introduction of a feature called Abuse Filter is also to be studied, as it allows administrators the ability to craft pattern-based rules that will automatically perform certain actions when a matching edit is made. Most of the current rules simply inform the editor that the edit was automatically identified as potential vandalism and asks him or her to confirm the action. However, the system allows highly-customized rules: for example, a particular user who is generally productive but incendiary on Middle East topics can be denied the ability to edit all pages in the “Middle East” category. The implications of this new system are obviously staggering and present a rich field for further research.
I also suspect that many other websites which are considered part of “Web 2.0” – social news sites like Digg or reddit, media sharing sites like Youtube or Flickr, social network sites like Facebook or Twitter, and blogs – are held together by the work of software agents outside of the official codebase. For example, ‘the blogosphere’ can only exist as an entity capable of being summarized in a headline such as “the blogosphere is teeming with talk of Twitter” because of sites like Technorati, which actively works to demarcate and associate the web, associating links between blogs in such a way that the popularity of each can be operationalized and emergent trends identified. Like vandal fighting in Wikipedia, this process does not require the active coordination of the participants involved – but only because a significant amount is being performed in order to stitch together the myriad of blogs on the Internet into something like ‘the blogosphere.’ Further research should interrogate the way in which phenomena traditionally explained by referencing sociological macro-actors or ‘the wisdom of crowds’ can instead be attributed to the work of largely unofficial technological actors.
Appendix A. Bibliography


Braun, Simone, and Andreas Schmidt. 2007. Wikis as a Technology Fostering Knowledge Maturing: What we can learn from Wikipedia. In *7th International Conference on Knowledge Management (IKNOW '07)*. http://publications.professional-learning.eu/braun_schmidt_WikisForKnowledgeMaturing_IKNOW07.pdf.


Cosley, Dan, Dan Frankowski, Loren Terveen, and John Riedl. 2007. SuggestBot: using intelligent task routing to help people find work in wikipedia. In


Kuznetsov, Stacey. 2006. Motivations of contributors to Wikipedia. SIGCAS 194


http://features.slashdot.org/article.pl?sid=05/04/18/164213.


Wray, K. Brad. 2009. The Epistemic Cultures of Science and Wikipedia: A

