The Work of Sustaining Order in Wikipedia: The Banning of a Vandal

R. Stuart Geiger **David Ribes**
Communication, Culture, and Technology Program
Georgetown University
3520 Prospect St NW, Suite 311, Washington, DC 20057 USA
{rsg33, dr273}@georgetown.edu

ABSTRACT
In this paper, we examine the social roles of software tools in the English-language Wikipedia, specifically focusing on autonomous editing programs and assisted editing tools. This qualitative research builds on recent research in which we quantitatively demonstrate the growing prevalence of such software in recent years. Using trace ethnography, we show how these often-unofficial technologies have fundamentally transformed the nature of editing and administration in Wikipedia. Specifically, we analyze 'vandal fighting' as an epistemic process of distributed cognition, highlighting the role of non-human actors in enabling a decentralized activity of collective intelligence. In all, this case shows that software programs are used for more than enforcing policies and standards. These tools enable coordinated yet decentralized action, independent of the specific norms currently in force.

Author Keywords
Wikipedia, wiki, bots, collaboration, distributed cognition, ethnography, social, qualitative, trace ethnography, vandalism

ACM Classification Keywords
H.5.3 [Information Interfaces]: Group and Organization Interfaces – Collaborative computing, Computer-supported cooperative work, Web-based interaction, H.3.5 [Information Storage and Retrieval]: Online Information Systems, K.4.3 [Computers and Society]: Organizational Impacts – Computer-supported collaborative work

General Terms
Human Factors, management

INTRODUCTION

From 21:20 to 21:31 on 19 February 2009, an unregistered (anonymous) user of Wikipedia made five edits to the article "Before we Self Destruct," a then-unreleased music album. The edits made were to the track listing and guests' section, swapping and replacing a significant number of titles and guests, including one that removed the track "Do What It Do" and added "Munch On My Penis" in its place. This edit, having triggered various vandalism-detection algorithms, appeared to a large and diverse set of users – human and non-human – who were monitoring changes to Wikipedia in near real-time using various semi- and fully-automated tools.

Six minutes later, at 21:37, a Wikipedia editor using one such program (called Huggle) advanced his queue and was shown the user's edit for the article on "Before we Self Destruct," along with contextual information about the anonymous user's previous edits. This editor was presented with a 'diff' – or a side-by-side comparison of changes that renders easily visible the text that had been changed in the edit (see figure 2). On his screen, Huggle also let the editor know that there were four previous edits made by this user to the article in the past few minutes. With the click a single red button, he removed all five of the anonymous user's edits to the article, reverting it to the condition left by the previous editor. This was his twentieth revert that day, and he would go on to make over 180 more over the next four hours. The previous minute, he had reverted edits from anonymous users on three articles, and in the next minute he would go on to revert edits from two articles.

Scholarly and popular accounts of Wikipedia, the self-proclaimed "free encyclopedia anyone can edit," often wonder at its near-immunity to vandals and spammers. They tend to posit a staggering number of insomniac reviewers and assume that volunteers must be constantly reverting and blocking or banning malevolent users, keeping the project from degenerating into anarchy. While such a view is partially correct, it ignores the heterogeneous assemblages of human and non-human actors deployed in the identification and temporary blocking of malicious contributors – a rather routine activity that occurs hundreds of times each day.

In this paper, we examine the process of counter-vandalism in Wikipedia, detailing the way in which participants and their assisted editing tools review contributions to Wikipedia and enforce various normative and epistemological standards. Such 'vandal fighters' have been identified as numerous and "organizationally important" [34], serving as the encyclopedia's first line of defense. They are also many newcomers' first introduction to the encyclopedia project's policies, standards, and procedures. Fully-automated anti-vandalism bots, a key non-human actor in this process, have

CSCW 2010, February 6–10, 2010, Savannah, Georgia, USA.

also been theorized by other researchers as being critical in stemming the rising tide of vandalism in Wikipedia [21].

Using the extant data Wikipedia automatically keeps on all edits, we trace in detail the blocking of the vandalous user introduced above. We focus on this anonymous user's edits and the process which led to that user's eventual blocking from Wikipedia after making approximately twenty inappropriate edits in a one hour period. The edits made were identified as vandalism and reverted by many different editors with many different tools and mechanisms to coordinate their work within Wikipedia. From autonomous software agents and semi-automated programs to user interface enhancements and visualization tools, these actors actively reshape the way in which editors engage with Wikipedia and its content. Together, they make possible a kind of epistemological enforcement that often requires little to no specific knowledge about a given article.

Contrary to common opinion, we show that the process of editing in Wikipedia is not a disconnected activity in which atomistic editors enforce their view of the world on others. Vandal fighting is instead shown to be a process of distributed cognition, through which users come to know their project and the users who edit it in a way that would otherwise be impossible for a single individual. Drawing from the work of Ed Hutchins, we claim 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. These semi- and fully-automated tools constitute an information infrastructure that makes possible the quick and seamless processes of valuation, negotiation, and administration between countless editors and issues.

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 social structures. While there is no doubt that the project's shared norms, codified standards, administrative processes, and formal institutions hold together the social order through which encyclopedia articles are produced and negotiated, we argue that such purely social actors are not sufficient explanations for the functioning of Wikipedia. The encyclopedia project's unlikely and unexpected success must also be attributed to a whole host of technological actors, who diligently work alongside human editors in the editorial and administrative process. Such human and non-human actors collectively but contingently comprise ad-hoc vandal-fighting networks, and this fast-paced, highly-mediated mode of distributed cognition is the way in which many Wikipedians come to apprehend and know their project, its content, and those who edit it. These humans and non-humans work to produce and maintain a social order that makes possible the collaborative production of an encyclopedia with hundreds of thousands of diverse and often unorganized contributors.

WIKIPEDIA AS A PROBLEM OF ORDER

One key research question asked by scholars apropos Wikipedia is how such a project could exhibit such stability and quality, given that nearly anyone with an Internet connection has the capacity to edit nearly any article as they see fit. While the Wikipedia community of editors is a highly technologically-mediated group, many of the explanations given by behavioral, informational, and even computational scientists have limited themselves to the level of "social explanations". In other words, these scholars have taken social forces and structures (norms, procedures, standards, cultural mores, governance institutions, discourses, power relations, roles, etc.) to be the source of the project's stability. While there are some articles that discuss the role of the software – particularly the 'anyone can edit' feature – in enabling the processes through which users produce and maintain social order [33,31,19,11,10], few have focused on the how largely-unofficial software is used by Wikipedians.

A partial explanation for why most social scientific research into Wikipedia has paid insufficient attention to these technological tools may be because of findings drawn from data collected in 2006. These figures showed that at their highest levels, such tools only comprise about 2 to 4 percent of all edits to the site [14], and they were largely involved in single-use tasks like importing public domain material [27]. As such, these unofficial tools have been implicitly characterized in the literature as mere force-multipliers, increasing the speed with which editors perform their work while generally leaving untouched the nature of the tasks themselves. Because of this, social research about Wikipedia has largely focused on unraveling the standards and practices through which editors coordinate and negotiate. For example, studies of Wikipedia's "policy environment" [2] or various designated discussion spaces have operated on this human-centered principle, demonstrating the complex and often "bureaucratic" [5] procedures necessary for the smooth-functioning of the project.

Most articles discussing technological tools in Wikipedia which explicitly perform some social function or are discussed as having some effect on the sociality of the project are proposing new tools [32,4,35,8,9], few of which have been taken up by the Wikipedian community. 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 tend to propose new algorithms [20,26,1] or evaluate the effectiveness of existing implementations [21]. In all, existing research has largely focused on the ways in which technology in Wikipedia has made the editorial process more efficient, transparent, and effective. While early research into the Wikipedian community provided rich accounts of the ways in which participants used various elements of the software [32,3], such work is outdated given the rise of new bots and editing tools. There has not been 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.

Bots, Scripts, and Other Tools

Bots – short for 'robots' – are fully-automated software agents that perform algorithmically-defined tasks involved with editing, maintenance, and administration in Wikipedia. For example, the first notable bot in the project (RamBot) imported public domain census data into articles about cities and towns. Other early bots trawled through articles, fixing simple grammatical or stylistic errors – like capitalizing certain unique proper nouns. At present, some of the most active bots are those that review every edit made in real time, using sophisticated heuristics to revert blatant incidents of spam and vandalism. However, there exist many different kinds of bots. While earlier research [14] showed that bots only made 2 to 4 percent of all edits in 2006, we have previously found [12] that this number has grown dramatically: at present, bots make 16.33% of all edits.

In addition, our data collection has allowed us to identify the prevalence of a new kind of technological tool which has emerged on the scene in Wikipedia: assisted editing programs. The traditional method of editing wiki pages using the wiki software is to review them, click the "edit this page" button, make whatever changes are deemed necessary in a text box, and then click submit. Assisted editing programs significantly alter editing work by automating various elements of the process, making the process faster and more efficient; in addition, such tools enable a distributed form of cognition among Wikipedia's decentralized editorial base.

One class of programs allows users to view all edits made to Wikipedia in a real time queue, 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 added commonly misspelled words, telltale signs of vandalism, or those made by anonymous users – among many other criteria. Our research has also shown the growing use of these programs since their emergence in late 2006: as of 2009, over 12 percent of all edits to the project are made with assisted tools. In some pages which are used to coordinate administrative work within Wikipedia, this figure is significantly higher. On one such page – Administrator Intervention against Vandalism (AIV), a noticeboard for suspected vandals – this figure is as high as 75 percent and has grown substantially since 2006 (Figure 1) [12].

Trace ethnography: a method for studying distributed cognition in sociotechnical networks

In order to explore the distributed action of vandal fighting, this research has made use of trace ethnography, a novel method for studying the complex interactions that occur in sociotechnical systems. In this section, we introduce the method as a way of studying the seemingly ad-hoc assemblage of editors, administrators, bots, assisted editing tools, and others who constitute Wikipedia's vandal fighting network. At its core, trace ethnography is a way of generating rich accounts of interaction by combining a fine grained analysis of the various 'traces' that are automatically recorded by the project's software alongside an ethnographically-derived understanding of the tools, techniques, practices, and procedures that generate such traces. For our case, one of us has been an active editor in the English-language Wikipedia for many years, and has spent a significant amount of time as a vandal fighter, using and interacting with many different kinds of tools and bots. We have therefore arrived at a rich understanding not only of the human work involved in the blocking of a vandal, but also of the role of the technological infrastructure. We have then proceeded to systematically reassemble markers and logs of various activities performed by human and non-human actors within Wikipedia.

This kind of methodology is made possible by the rich amount of publicly-available data regarding particular edits to Wikipedia. Like many version control systems (VCS) used to maintain software code, the MediaWiki software platform upon which Wikipedia runs automatically preserves a copy of each revision, along with metadata such as the editor's username or IP address, the time the edit was made, and a comment field where editors can give a short summary of their edit. These publicly-available revision histories can be generated for articles as well as users, allowing us to trace the edits of a particular anonymous vandal and then the subsequent edits made to the pages they vandalized. These edit summaries comprise the bulk of the traces that we used to generate our descriptions actions and interactions of the various actors described in our case study.

By default, most assisted editing tools preface or append edit summaries with a short, unique marker identifying that the edit was made with the tool in question – (HG) for Huggle, (TW) for Twinkle, and so on. In performing different highly-

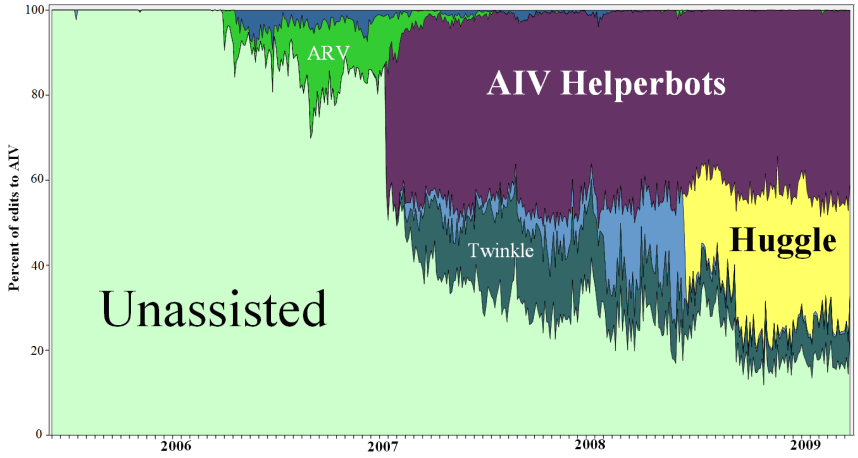

Figure 1: Edits to Administrator Intervention against Vandalism by tool [12]

specialized actions, these tools also generate standard edit summaries that allowed us to know which actions were taken; for example, when rolling back multiple vandalous edits, Twinkle makes an edit summary such as: "Reverted 5 edits by 72.68.228.176 identified as vandalism to last revision by Alansohn. (TW)." Combined with ethnographic knowledge regarding the capabilities of these editing tools and the kind of edit summaries they make, we were able to reconstruct the actions of editors as they went about banning a user: the software they used, the evidence they were presented with, and even the buttons they clicked. From this, we can also give rich accounts of the roles of the assisted editing tools themselves as they automatically wrote templated warnings to a user's talk pages – public wiki pages created by the software system for each user.

Our reconstruction of users', editors', and tools' coordinated actions using trace ethnography provides a detailed qualitative description of the human and non-human work which led to the banning of a vandal. Taking ethnography to be the generation of ethno-graphs – literally, 'the people in writing' –our method provides notable advantages over single or even multi-site ethnography, as it allows us to capture network-level phenomena. It would be rather difficult for an ethnographer at a single site to give the kind of full and detailed accounts of human-computer interaction at the network level without relying on an analysis of traces. As such, trace ethnography extends a long line of qualitative research of distributed collaborative work in technologically-mediated communities, especially studies of bug tracking [22,25], open source software development [23,24], and repair technicians [18]. While our methodology is also similar in spirit to Lucy Suchman's landmark studies of human-machine interactions [29], her use of videotapes to capture the actions of users and technologies would not easily scale to the level of Wikipedia's vandal fighting network. Using our method, the collective work of banning a vandal is rendered *directly observable* by following the traces left in Wikipedia.

Trace ethnography is also heavily influenced from other forms of ethnography, most notably Diane Vaughn's historical ethnography of the Columbia shuttle explosion [30]. Vaughn was able to assemble a narrative using the thick documentary evidence kept by NASA and various subcontractors, tracing out the various layered interactions that led to the decision to launch the shuttle. On a more practical level, our methodology also draws significantly from Bruno Latour's work on circulating references in science. For example, in an ethnographic essay on fieldwork [17], Latour traces out the cascading chains of data analysis that ultimately turn acres of forest and savannah into a crisp scientific chart. Furthermore, we also take from his more theoretical work on delegation as a heuristic for symmetrically analyzing the way in which both humans and machines contribute to the production of social order. This literature base provides a foundation upon which our empirical work of trace ethnography can provide rich theoretical accounts of action and practice.

Theorizing Vandal Fighting as Distributed Cognition

In *Cognition in the Wild* [13], informed by ethnographic research on board a U.S. Navy ship, Ed Hutchins tells of the astounding amount of informational and cognitive work must be performed in order to keep the ship on course at any given time. In order to cope with these demands, information gathering and processing is distributed to crew members, who regularly collect data, analyze it, and pass the results to others. Hutchins' research directly opposes that of cognitive scientists and others who believe that cognition occurs solely in the heads 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 first glance at something like a Navy ship (or Wikipedia) may give the illusion of natural regularity, but Hutchins repeatedly emphasizes the sociality of such cognitive systems.

In his anthropological accounts, Hutchins goes into rich 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. 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. Such systems are powerful and well-functioning even when their members are seemingly in cognitive imbalance or isolation.

As with Hutchins' navigational charts, the technological actors in Wikipedia such as Huggle make what would be a difficult task into a mundane affair. As we will see, reverting an edit becomes a matter of pressing a button, and blocking a vandal clicking a 'yes' button in a dialog box. However, there are key differences between the kind of distributed cognition at work in a largely professionalized US Navy ship and an all-volunteer Wikipedian vandal fighting network. As such, this account helps shed some light onto the open mystery as to how a group of diverse, uncoordinated, and often un-credentialed individuals can come to collectively

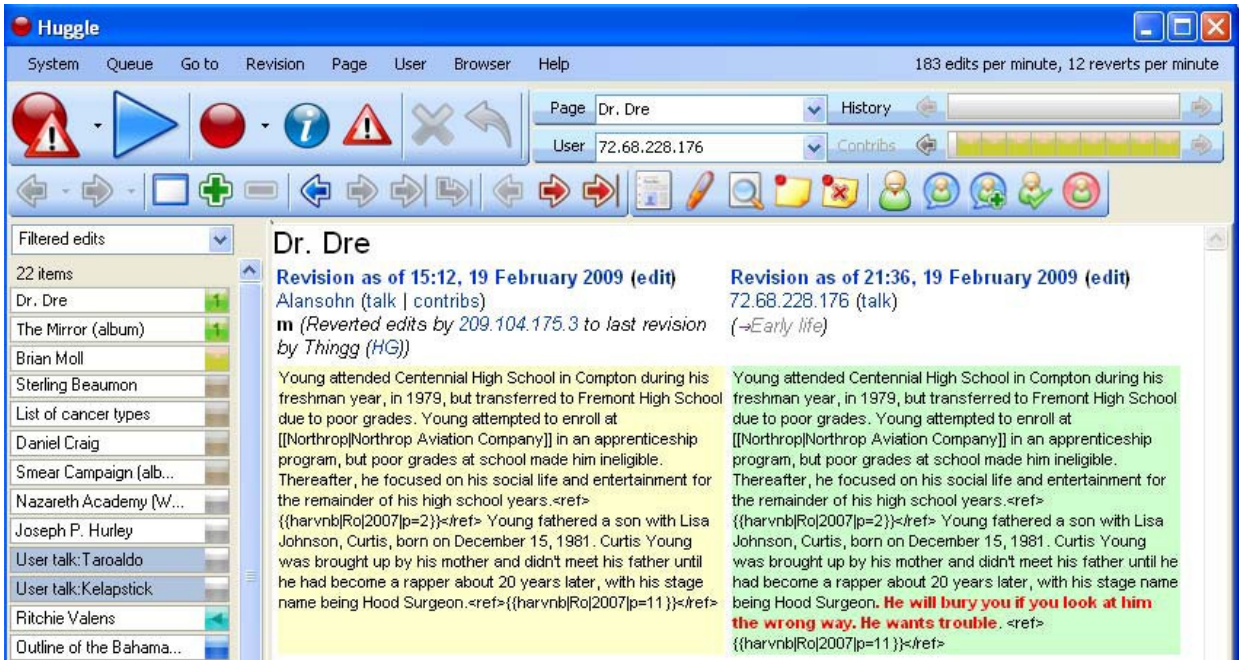

Figure 2: A diff presented in the Huggle assisted editing tool, viewing an edit made by the anonymous user (simulated)

build and maintain the world's largest and most public encyclopedic reference work.

CASE STUDY: THE BANNING OF A VANDAL

While it is clear that bots and assisted editing programs comprise a large portion of all edits made to Wikipedia, their impact on the project can only be understood in their use. In order to show tool use in context, in this section we provide an account of the banning of a vandal, paying close attention to the technological tools through which this process was made possible. While previous research has reviewed the largely social mechanisms involved with formal [33] and informal [28,11,15] review processes, little research has been performed in Wikipedia's more immediate forms of review and revision. In large part to a vast array of interoperable tools, bots, and standards, the process of vandal fighting is becoming increasingly automated.

Huggle (Figure 2), which is the most widely-used assisted editing program, transforms the vandal fighting process by dramatically enhancing the way in which users interact within Wikipedia. In this stand-alone program, edits are contextually presented in queues as they are made, and the user can perform a variety of actions (including revert and warn) with a single click. The software's built-in queuing mechanism, which by default ranks edits according to a set of vandalism-identification algorithms, is a form of technological delegation *par excellance*. Users of Huggle's automatic ranking mechanisms do not have to 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. For example, in the default 'filtered' queue, edits that contain a significant removal of content are placed higher; those that completely replace a page with blank text are even marked in the queue with a red 'X'. The queue is also ranked by the kind of user who made the edit: anonymous users are viewed as more suspicious than registered users, and edits by bots and Huggle users are not even viewed at all. Users whose edits have been previously reverted by a number of assisted users are viewed as even more suspicious, and those who have been left warnings on their user talk page (a process explained below) are systematically sent to the top of the queue.

Another key feature of Huggle is the way in which users can, upon reverting an edit as vandalism, automatically leave a warning for the offending editor. This mechanism makes use of user talk pages, which are public wiki pages created by the software system for each user. In the process, vandal fighters leave pre-written, templated warnings on the user talk pages of offending editors, ostensibly to notify him or her that the edit in question was not encyclopedic. However, because of their public nature, user talk page messages have become a kind of database, cataloging identified incidents of vandalism for particular users. It is of note that there are hundreds of these warnings, which vary widely, but almost all are categorized into four levels of increasing severity and tone. As such, Huggle and other programs can determine a user's previous record of vandalous edits by retrieving the severity level of previous warnings (if any) on his or her talk page.

Many human editors may be involved in the identification and reporting of a vandal, and user talk page warnings network the cognitive work of vandal identification for other vandal fighters. As the pre-written warning messages are

divided into four levels of severity, it is standard practice to issue a first level warning for the first incident of vandalism and escalate through the chain as more incidents are identified. Generally, administrators will not temporarily block users from editing if they have not received four warnings. Because of this, the practice of warning operationalizes each offending edit into the social structure through which administrators and editors come to know users as vandals. The work performed by many distinct vandal fighters can be collated and then compressed into a single number, visible to a wide array of human and non-human actors. For example, Huggle and a few vandalism-reverting bots can review warnings left for a user and adjust their actions accordingly, even notifying administrators when a user with a fourth-level warning is reverted and warned.

While Huggle is the most prevalent assisted editing tool, others are in high use as well. The next most popular tool is Twinkle, which is a user interface extension that runs inside of a standard web browser. Twinkle adds contextual links to pages in Wikipedia allowing editors to perform complex tasks with the click of a button – such as rolling back multiple edits by a single user, reporting a problematic user to administrators, nominating an article for deletion, and temporarily blocking a user (for administrators only). Other tools include Lupin's anti-vandal tool, which provides a real-time in-browser feed of edits made matching certain algorithms (such as obscene words or commonly misspelled words), allowing users to review and correct such errors at their discretion. Of note is the fact that these tools are largely unofficial and maintained by members of the Wikipedia community.

A Vandal Emerges

In the case presented in the introduction, a series of edits to an article were reverted within minutes by a vandal fighter using the Huggle tool. However, this was only the beginning of the story of how this user came to be banned as a vandal. The anonymous user who made the edits was not deterred by the reversion of edits or warnings and continued to vandalize other articles. The vandal's actions had not yet been sufficient to warrant a block by Wikipedia's administrators. Blocking could not occur following the first incident of vandalism; it required a network of decentralized vandal fighters, each making separate determinations, which then circulated through a complex distributed chain before an administrator finally determined that the user was indeed a vandal worthy of a temporary ban. In this section, we trace the actions of the user, a bot, and several editors using assisted editing tools, immediately following our introductory vignette.

We left off at 21:37, as a Wikipedian editor using Huggle reverted an offending edit to an article for the album "Before we Self Destruct". Yet unbeknownst to this editor, the anonymous user had vandalized another article minutes after the first incident, this time to the article for the album "The Mirror". Also at 21:37, a different Wikipedian editor reviewed the 'diff' of the edit – a before-and-after comparison that is built into the MediaWiki software. The user had also installed Twinkle, which inserted a link titled "[rollback]" into this page. When this link was clicked, Twinkle set into motion a pre-scripted path of action that first reverted the edit in question and then popped up both the vandal's user talk page and a specialized dialog box. Here, the vandal fighter was presented with hundreds of pre-written messages, each divided into four categories of severity. With no recent warnings left on the vandal's talk page, the vandal fighter was satisfied with Twinkle's default: a politely worded 'first-level' message that "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." This template, coded as "uw-vandalism1", also informs of a "welcome page" if they wish to help editing "constructively to this encyclopedia."

However, it does not appear that the vandal was interested in this goal, as at 21:43, another edit was made to an article about an album, "808s & Heartbreak", this time removing an entire section. This edit was placed into the queues of many Huggle users, as the software prioritizes mass removal of content by anonymous users who have vandalism warnings left for them. In fact, a green "1" appeared next to the article's name in the edit queue, indicating that a first-level warning had been issued. Advancing to this edit, the anonymous user's edit appeared on the first vandal fighter's screen, who had initially reverted the user's edits without warning. However, this time, he clicked a different red button that simultaneously reverted the edit and left a pre-formatted message on the anonymous user's talk page. In performing this action, the Huggle program examined the user's talk page and found the warning that the second vandal fighter issued with Twinkle. Because of Huggle's automated ability to uncover warnings made by other editors, it automatically issued a warning that was slightly stronger in tone than the previous first-level comment.

The user was still not dissuaded, making another edit to the "808s & Heartbreak" article by 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. In a matter of seconds, a bot named ClueBot examined this edit and the contribution history of the anonymous user, finding it to be a clear-cut case of vandalism. The bot reverted the edit in seconds – before any other human or non-human vandal fighter was able to react – and then moved to the anonymous user's talk page. It received a list of all messages left for the user, identified the previous message left as a second-level warning, and issued a third-level warning, asking the user to "Please stop. If you continue to vandalize Wikipedia, as you did to 808s & Heartbreak, you *will* be blocked from editing." This task completed, ClueBot moved to another edit; this was the 592,829th edit that ClueBot had reverted since it had begun operation in August of 2007.

However, the anonymous user was not finished, adding the phrase "KanGay west is the proud owner of a vagina" to the

end of the same section. A minute later, a third vandal fighter advanced his Huggle queue, found this edit, and clicked the same red 'revert and warn' button. The Huggle program, seeing the bot's third-level warning, automatically issued a fourth-level warning, which presented the user with an ultimatum; with the template code "uw-huggle4", this pre-written message told the vandal that "You will be blocked from editing the next time you vandalize a page, as you did with this edit to '808s & Heartbreak.'" As may be expected, this final warning was ignored as the anonymous user edited the "808s & Heartbreak" article within a minute, changing the length of the track from "52:05" to "52:05FUck."

This edit appeared on the top of many Huggle queues with a deep red "4" icon – indicating that this user had received a fourth-level warning and was a highly-likely suspect for vandalism. The third vandal fighter caught this user's edit again, clicking the same button to revert this edit and warn the user as he did for the previous edit. However, when the Huggle software examined the user's talk page, it found that the previous message was a level four warning; as such, it asked the editor if he wanted to report the user as a vandal. He clicked the "Yes" button, and the Huggle software made an edit to a different page: "Wikipedia: Administrator intervention against vandalism" or AIV, as it is known.

The AIV page (Figure 3) is used by vandal fighters to make formal ban requests to administrators who alone have the technical and social authority to issue temporary blocks. Unlike many of the 'meta' pages in Wikipedia, the page has become more of a queuing mechanism than a discussion forum. In the language of actor-network theory, AIV is the obligatory passage point [6] between vandal fighters and administrators, serving as a clearinghouse to facilitate the quick and centralized processing of ban requests. In reporting the anonymous user to AIV, the Huggle program collected three edits which had been marked as vandalism in the previously-issued warnings. Any user who visited the AIV page could click any of these links to see a permanent hyperlink to a diff of an edit. The diff would contextualize the edit, just as the original vandal fighters had seen and based their decisions upon. The report also contained auto-generated links to facilitate specific vandal-fighting tasks.

By now, a vandal fighting network had fully assembled from the vast collective of Wikipedia to combat the increasingly frantic edits made by the anonymous user. Both humans and software had become highly-tuned to this user's actions, intensely watching his or her actions and reverting each vandalous edit in seconds. Before the AIV report was formally reviewed by an administrator monitoring the noticeboard, the anonymous user made another edit, this time to the article "Conan O'Brian." In this edit, he replaced the occupations of Conan's parents with "pizza delivery man" and "majure in flamingos." In a single click, the second vandal fighter – who also happened to be an administrator, technically but not socially able to block the vandal at any time he wished – issued the 'revert and warn' command, as he had done once before. The Huggle software took note of the fact that a report existed for this user at AIV, and asked the administrator if he wished to issue a temporary block. He did, and so the software prompted him for a summary justification, a block length, and various options. Choosing the defaults ("Vandalism" and for 48 hours), Huggle issued a command to the MediaWiki software that placed the anonymous user on a list of blocked users and automatically issued a corresponding message on the user's talk page.

In the seconds between when the administrator reverted the vandalism to the Conan O'Brian article and submitted the command to block, the vandal was able to make one final edit - repeating the same edit to the Conan O'Brian article. The third vandal fighter caught this edit using Huggle, and, unaware that the administrator was blocking the user at that very instant, instructed the program to again revert and warn. Yet with four warnings and an active report at AIV, there was nothing else Huggle could do in the name of this non-administrator except append this incident of vandalism to his original report, further attempting to enroll a willing administrator into the ad-hoc vandal fighting network. This was unnecessary, and seconds later, the software finished processing the administrative block. The next minute, a different bot – "HBC AIV helperbot7" – automatically removed the third vandal fighter's now-obsolete report.

DISCUSSION

A significant number of actors are required to act in coordination with each other to ban a vandal. As this case shows, four human editors and one bot each made separate judgments of vandalism within a fifteen minute period, which ordinarily would not be sufficient to make such a determination. Yet through the specific software used by the editors, identified incidences of vandalism were reported to the user's talk page, which was more of database for other vandal fighters than a space for dialogue with the anonymous editor. This illustrates that vandal fighters are not merely individually assisted by such tools, but rather are joined together by the various software programs into a decentralized network. While each editor made local judgments as to the veracity or appropriateness of specific contextualized edits, they collectively came to identify users who were problematic and thus deserving of a temporary ban. The cognitive work of identifying and banning a single user was distributed across this heterogeneous network. This redistribution of work should also be seen as a transformation of the moral order of Wikipedia, changing the very methods by which edits are evaluated, content is reverted and users are banned. This redistribution of moral agency to automated and semi-automated tools has significant consequences how vandal fighting and editing work is performed.

Redistributing the work of editing

The most notable feature of such assisted editing programs is the speed and efficiency afforded to busy vandal fighters, however they do much more than this. These tools greatly

lower certain barriers to participation and render editing activity into work that can be performed by 'average volunteers' who may have little to no knowledge of the content of the article at hand. Such a reviewing process is in stark contrast to the more traditional forms of professional and academic knowledge production by experts who are able to contribute by virtue of their knowledge of a domain. The domain expertise of vandal fighters is in the use of the assisted editing tools themselves, and the kinds of commonsensical judgment those tools enable.

Like Hutchins's analysis of navigational charts, technological actors like Huggle make what would be an involved evaluative process into a mundane affair. That is, a significant portion of vandalism is rendered clearly identifiable as such, even if one knows nothing of the topic at hand, e.g., in a page on a music album, the letters "FUck" should simply not appear in a field for track length. Likewise, appending a single sentence to the end of a quote is immediately visible and inherently suspicious because of the diff; such an edit would be very difficult to identify if a reviewer were simply reading the article from beginning to end, in particular if the reviewer was unfamiliar with topic at hand. It is the tools themselves that re-present change in ways that render them visible to any editor, and enable a common sense judgment about whether it is vandalism.

The most obvious cases of such vandalisms 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: when we presented edits using the same kinds of tools and programs that vandal fighters use in their daily work, the edits in question were rendered visibly suspicious because they were displayed in such a manner. While nonsense or obscenity may be easy to spot when proofreading the entire article, removal of entire sections is a common form of vandalism that is difficult to detect by merely reading the article. 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 mechanism 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 a sentence.

The Huggle program's queuing mechanism is another way in which edits are further transformed, contextualized, and abstracted. Ranking edits according to a pre-established set of vandalism-identification algorithms and heuristics, the queue is a form of *delegated cognition*. Users of Huggle's automatic ranking mechanisms do not generally decide on their own 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. Ideally, the vandal fighter reserves his or her cognitive duties for those which cannot be replicated by computerized algorithms. This means that the editorial process can leverage the skills of volunteers who may not be qualified to formally review an article in, for instance, an academic peer-review setting, but can contribute to the maintenance of order. This capability is similar to Collins and Evans' [7] distinction between interactional and contributory expertise: one does not need to have the technical, literary, or academic skills or motivations to author an article in order to patrol it.

Yet the most notable aspect of the Huggle software – as well as Twinkle, ARV, AIVer, and other assisted editing tools – is the way in which they collectively enable a form of distributed coordination among otherwise disconnected vandal fighters. As was made apparent, a significant number of human actors are required to act in concert with each other to ban a vandal. In this case, four editors – three humans and one bot – each made separate determinations of vandalism within a fifteen minute period. Through the specific software used by the editors, identified incidences of vandalism were reported to the user's talk page. This shows that the process of warning is not only to inform a user that their actions are disrespectful or unwanted, but is also an act of coordination, largely conducted by semi-automated software.

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 is abstracted and contextualized by incorporating it into a warning template. Through the user talk page warning, a record of vandalism is created that can be subsequently deployed by any vandal fighter, administrator, or other interested user. In the language of distributed cognition or actor-network theory, this is a form of immutable mobile inscription. 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 simply have the software automatically incorporate this information into its decision-making process. With various programs and user interface extensions, an editor can quickly determine if a user has been sufficiently warned by others in vandal fighting community – regardless of the tools they used in the process – and report those who continue to abuse their editorial privileges. This illustrates the two-way nature of this semi-automated and distributed system of cognition, as the incidents identified by other vandal fighters can be captured and systematically deployed as evidence in other spaces.

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 site. 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. They constitute a largely invisible infrastructure that has been increasingly critical in insulating Wikipedia from vandals, spammers, and other malevolent editors.

Redistributing Moral Agency

One of the most striking elements of Wikipedia's vandal fighting networks is the extent to which it transforms the decision making process in reviewing edits. In a setting such as Wikipedia, such decisions are key turning points in deciding what is valid or invalid content and who are the legitimate or illegitimate contributors to a base of knowledge. Such acts of inclusion and exclusion may be necessary, but they are inherently moral in quality, speaking to questions of who is left out and what knowledge is erased. Such transformations to the participants and process of decision-making demands closer scrutiny from many angles. It is for this reason that the argument that bots and assisted editing tools are merely force multipliers is narrow and dangerous: proponents of such an argument see only the speeding up of an existing process, rather than its transformation.

For example, as we have discussed, the Huggle software has a pre-defined a set of criteria for identifying likely vandalisms. In certain cases, such as with ClueBot, bots automatically revert edits according to criteria such as obscenity, patent nonsense, mass removal of content, and various metrics regarding the user who made the edit. It is this last aspect that is most significant in terms of morality, as the detection algorithms explicitly discriminate against anonymous and newly-registered editors. In addition, this system enables a new kind of moral order in Wikipedia, making it possible for editors and administrators to track the extent to which a certain user or IP address has vandalized through talk page warnings. If this has made the process more participatory, it is precisely because it has become more automated and inflexible. In and outside of the Wikipedian community, tools like Huggle are often compared with video games in both serious critiques and humorous commentaries: reviewing changes are presented in easily comprehensible and attractive graphics; reverting an edit is a matter of clicking a button.

We should not fall into the trap of speaking of bots and assisted editing tools as constraining the moral agency of editors. Rather, it is that the delegation of certain tasks to these tools makes certain pathways of action easier for vandal fighters and others harder. For example, it possible in Huggle to circumvent the standardized pathway of four sequentially-escalating warnings described above: if an incident of vandalism is particularly egregious, a vandal fighter can issue a fourth-level warning when the software would have automatically given a much lower level. Similarly, users can reconfigure their queues to not view anonymous edits as more suspicious, or even to only review edits made by Huggle users. While these and many other workarounds are possible, they require a greater effort and a certain technical savvy on the part of their users. As Bruno Latour notes, "In spite of the constant weeping of moralists, no human is as relentlessly moral as a machine, especially if it is as 'user friendly' as my computer." [16] Ultimately, these tools take their users through standardized scripts of action in which it always possible to act otherwise, but such deviations demand inventiveness and time.

CONCLUSION

As this case shows, technological tools like bots and assisted editing programs have a significant social effect on the kinds of activities that are made possible in Wikipedia. While this process was facilitated through the construction of various social artifacts, such as templated warning messages and codified standards of vandalism, the social roles of technological artifacts is difficult to ignore. Without knowing of such non-human actors at work, it may seem unfathomable that such coordination against vandals could even be possible, even given the social infrastructures detailed by previous researchers. Yet in light of these infrastructural assemblages through which Wikipedia's standards and policies are enforced on a daily basis, such acts of enforcement seem far less spontaneous or mystical. In future research, bots must be examined as more than mere force-multipliers or irrelevant users. Bots can reshape the social world by enabling a specialized discursive space for the coordination of vandal fighting tasks. In addition, assisted editing programs must also be studied for their social effects, given the way in which they were shown to automatically operationalize normative enforcement. This is a particularly interesting opportunity for study, especially regarding the way in which such tools transform the nature of user interaction.

Finally, this research has shown the salience of trace ethnography for the study of distributed sociotechnical systems. The method is best for revealing the often invisible infrastructure that underlie routinized activities, allowing researchers to generate highly-empirical accounts of network-level phenomena without having to be present at every node. Trace ethnography does have its limitations, as it does not allow researchers to grasp the larger sociocultural significance or history of the activities at hand. However, trace ethnography is fully compatible with other qualitative and quantitative methods, including traditional ethnographic, historical, archival, interview, survey, and statistical methods. For full account of sociotechnical phenomena, mixed method studies may be appropriate; if one wished to learn, for example, how such a network of vandal fighting humans and technologies was originally constituted, or how the system of four warnings came to be the standard for blocking vandals.

ACKNOWLEDGEMENTS

We would like to thank Jed Brubaker, Margarita Rayzberg, the CSCW reviewers, and countless Wikipedian editors.

WORKS CITED

1. Adler, B.T. and Alfaro, L.D. A content-driven reputation system for the Wikipedia. Proceedings of the 16th international conference on World Wide Web, ACM (2007), 261-270.

2. Beschastnikh, I., Kriplean, T., and McDonald, D. Wikipedian Self-Governance in Action: Motivating the Policy Lens. Proceedings of the Second International Conference on Weblogs and Social Media, (2008).
3. Bryant, S., Forte, A., and Bruckman, A. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. Proceedings of the 2005 conference on Supporting group work, (2005), 1-10.
4. Buriol, L.S., Castillo, C., Donato, D., Leonardi, S., and Millozzi, S. Temporal Analysis of the Wikigraph. Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, IEEE (2006), 45-51.
5. Butler, B., Joyce, E., and Pike, J. Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia. Proceeding of the 2008 SIGCHI conference on Human factors in computing systems, ACM (2008), 1101-1110.
6. Callon, M. Some elements of a sociology of translation: domestication of the scallops and the fishermen of StBrieuc Bay. In Power, Action and Belief: A New Sociology of Knowledge. Routledge & Kegan Paul, London, 1986, 196.
7. Collins, H. and Evans, R. The Third Wave of Science Studies: Studies of Expertise and Experience. Social Studies of Science 32, 2 (2002), 235-296.
8. Cosley, D., Frankowski, D., Terveen, L., and Riedl, J. SuggestBot: using intelligent task routing to help people find work in wikipedia. Proceedings of the 12th international conference on Intelligent user interfaces, ACM (2007), 32-41.
9. Demartini, G. Finding Experts Using Wikipedia. 2nd International ExpertFinder Workshop, (2007), 33–41.
10. Emigh, W. and Herring, S.C. Collaborative Authoring on the Web: A Genre Analysis of Online Encyclopedias. Proceedings of the 38th Annual Hawaii International Conference on System Science, IEEE (2005), 99.
11. Forte, A. and Bruckman, A. Scaling Consensus: Increasing Decentralization in Wikipedia Governance. Proceedings of the 41st Annual Hawaii International Conference on System Sciences, IEEE (2008), 157.
12. Geiger, R.S. The Social Roles of Bots and Assisted Editing Tools. Proceedings of the 2009 International Symposium on Wikis (Wikisym), ACM (2009).
13. Hutchins, E. Cognition in the Wild. The MIT Press, 1996.
14. Kittur, A., Pendleton, B., Suh, B., and Mytkowicz, T. Power of the Few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie. Proceedings of the 25th Annual ACM Conference on Human Factors in Computing Systems (CHI 2007), ACM (2007).
15. Kittur, A., Suh, B., Pendleton, B.A., and Chi, E.H. He says, she says: conflict and coordination in Wikipedia. Proceedings of CHI 2007, ACM (2007), 453-462.
16. Latour, B. Mixing Humans and Nonhumans Together. Social Problems 35, 3 (1988), 298-310.
17. Latour, B. Circulating Reference: Sampling the Soil in the Amazon Forest. In Pandora's Hope. Harvard University Press, Cambridge MA, 1999, 24.
18. Orr, J. Talking about machines : an ethnography of a modern job. ILR Press, Ithaca N.Y., 1996.
19. Pentzold, C. and Seidenglanz, S. Foucault@ Wiki: first steps towards a conceptual framework for the analysis of Wiki discourses. Proceedings of the 2006 international symposium on Wikis, ACM New York (2006), 59-68.
20. Potthast, M., Stein, B., and Gerling, R. Automatic Vandalism Detection in Wikipedia. In Advances in Information Retrieval. 2008, 663-668.
21. Priedhorsky, R., CHEN, J., Lam, S.T.K., Panciera, K., Terveen, L., and Riedl, J. Creating, destroying, and restoring value in wikipedia. Proceedings of the 2007 international Conference on supporting group work, ACM New York, NY, USA (2007), 259-268.
22. Reagle Jr, J. Bug Tracking Systems as Public Spheres. Techné,11,1 (2007) 32.
23. Sack, W., Détienne, F., Ducheneaut, N., Burkhardt, J.M., Mahendran, D., and Barcellini, F. A methodological framework for socio-cognitive analyses of collaborative design of open source software. Computer Supported Cooperative Work (CSCW) 15, 2 (2006), 229–250.
24. Scacchi, W. Free/open source software development. Proceedings of the 6th joint meeting of the European software engineering conference and the ACM SIGSOFT symposium on The foundations of software engineering, ACM (2007), 459-468.
25. Shukla, S. and Redmiles, D. Collaborative learning in a bug-tracking scenario. Conference on Computer Supported Cooperative Work, (1996).
26. Smets, K., Goethals, B., and Verdonk, B. Automatic Vandalism Detection in Wikipedia: Towards a Machine Learning Approach. In AAAI Workshop on Wikipedia and Artificial Intelligence, 43–48, AAAI, (2008).
27. Stvilia, B., Twidale, M., Smith, L., and Gasser, L. Assessing information quality of a community-based encyclopedia. Proceedings of ICIQ, (2005), 442-454.
28. Stvilia, B., Twidale, M.B., Smith, L.C., and Gasser, L. Information quality work organization in wikipedia. Journal of the American Society for Information Science and Technology 59, 6 (2008), 983-1001.
29. Suchman, L.A. Human-Machine Reconfigurations. Cambridge University Press, Cambridge, 2007.
30. Vaughan, D. The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA. University Of Chicago Press, 1997.
31. Viegas, F. The Visual Side of Wikipedia. Proceedings of the 40th Annual Hawaii International Conference on System Sciences, IEEE (2007).
32. Viegas, F., Wattenberg, M., and Dave, K. Studying Cooperation and Conflict between Authors with history flow Visualizations. In Proceedings of CHI, ACM (2004).
33. Viegas, F., Wattenberg, M., and McKeon, M. The Hidden Order of Wikipedia. In Online Communities and Social Computing. 2007, 445-454.
34. Welser, H., Kossinets, G., Marc, S., and Cosley, D. Finding Social Roles in Wikipedia. Paper presented at the annual meeting of the American Sociological Association, Boston, MA, AllAcademic, (2008).
35. Zeng, H., Alhossaini, M.A., Ding, L., Fikes, R., and McGuinness, D.L. Computing Trust from Revision History. Proceedings of the 2006 International Conference on Privacy, Security and Trust. ACM (2006).