NETWORK OF KNOWLEDGE: WIKIPEDIA AS A SOCIOTECHNICAL SYSTEM OF INTELLIGENCE

by

RANDALL M. LIVINGSTONE

A DISSERTATION

Presented to the School of Journalism and Communication
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

September 2012

DISSERTATION APPROVAL PAGE

Student: Randall M. Livingstone

Title: Network of Knowledge: Wikipedia as a Sociotechnical System of Intelligence

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the School of Journalism and Communication by:

Dr. Biswarup Sen	Chairperson
Dr. Janet Wasko	Member
Dr. Kim Sheehan	Member
Dr. Ryan Light	Outside Member

and

Kimberly Andrews Espy	Vice President for Research & Innovation/Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded September 2012

DISSERTATION ABSTRACT

Randall M. Livingstone

Doctor of Philosophy

School of Journalism and Communication

September 2012

Title: Network of Knowledge: Wikipedia as a Sociotechnical System of Intelligence

The purpose of this study was to explore the codependencies of the social and technical structures that yield Wikipedia the website and Wikipedia the community. In doing so, the research investigated the implications of such a sociotechnical system for the maintenance of the project and the emergence of collective intelligence. Using a theoretical framework informed by digital media studies, science and technology studies, and the political economy of communication, this study examined the material and ideological conditions in which Wikipedia has developed. The study's guiding research questions addressed the nature of Wikipedia's sociotechnical system and potential for collective intelligence, as well as the historical development of the project's technical infrastructure and the state of its technology-assisted collaboration.

A mainly qualitative multi-method research approach was employed, including document analysis, semi-structured interviewing, and social network analysis. A plethora of documents were carefully selected and examined to explore how and why decisions were made, policies implemented, and technologies adopted on the site. Additionally, 45 interviews were conducted with members of Wikipedia's technical community to understand the relationships between social and technical aspects of the project and the motivations of programmers who contribute automated tools. Finally, social network measures and visualizations were used to interrogate notions of collaboration and make more transparent the centrality of technology to the content creation process.

The study revealed that Wikipedia's technical development has been shaped by the dueling ideologies of the open-source software movement and postindustrial capitalism. Its sociotechnical system features the complex collaboration of human contributors, automated programs, social bureaucracy, and technical protocol, each of which conditions the existence and meaning of the others. In addition, the activity on Wikipedia fits established models of collective intelligence and suggests the emergence of a cyberculture, or culturally informed shared intelligence, unique to the digital media context. Software robots (bots) are central actors in this system and are explored in detail throughout this research.

CURRICULUM VITAE

NAME OF AUTHOR: Randall M. Livingstone

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
University of Massachusetts, Dartmouth
Bridgewater State College, Bridgewater, MA

DEGREES AWARDED:

Doctor of Philosophy, Communication and Society, 2012, University of Oregon
Master of Arts, Professional Communication, 2004, University of Massachusetts Dartmouth
Bachelor of Science, Computer Science and English, 2001, Bridgewater State College

AREAS OF SPECIAL INTEREST:

New/Digital Media
Communication Technologies
Political Economy of Communication
Advertising's Role in Society

PROFESSIONAL EXPERIENCE:

Graduate Teaching Fellow, School of Journalism and Communication, University of Oregon, 2008-2012

Assistant Director, Ballotti Learning Center, Suffolk University, 2005-2008

Senior Lecturer, Department of English, Suffolk University, 2004-2008

Teaching Associate, Department of English, University of Massachusetts Dartmouth, 2001-2004

GRANTS, AWARDS, AND HONORS:

Lorry I Lokey Scholarship, School of Journalism and Communication, University of Oregon, 2011

Arant Scholarship, School of Journalism and Communication, University of Oregon, 2010-2011.

Columbia Scholarship, School of Journalism and Communication, University of Oregon, 2009-2011.

Kappa Tau Alpha Honor Society, School of Journalism and Communication, University of Oregon, 2010

University Scholarship, Department of English, University of Massachusetts Dartmouth, 2001-2004

PUBLICATIONS:

Santana, A., Livingstone, R. M., & Cho, Y. (in press). Medium matters: Newsreaders' recall and engagement with online and print newspapers. *Newspaper Research Journal*.

Livingstone, R. M. (2011). Better at life stuff: Consumption, identity, and class in Apple's "Get a Mac" campaign. *Journal of Communication Inquiry, 35*(3), 210-234.

Livingstone, R. M. (2010). Let's leave the bias to the mainstream media: A Wikipedia community fighting for information neutrality. *M/C Journal, 13*(6).

ACKNOWLEDGMENTS

I would like to thank a number of people who have made this project possible, as well as those who have helped me tremendously throughout my graduate studies. First, a special thanks to my advisor and committee chair, Dr. Bish Sen, who has supported my interest in digital media over the past four years and inspired me to look deeply at its form, function, and impact. I am also greatly appreciative of my other committee members: Dr. Janet Wasko, Dr. Kim Sheehan, and Dr. Ryan Light. Thank you for all of the chats, all of the guidance, and all of the input over the years. Additionally, I would like to thank other faculty and staff at the University of Oregon who have helped me along the way, especially Dr. Leslie Steeves, Dr. Daniel Pope, Dr. Carl Bybee, and Petra Hagen.

This project was made possible by the openness and kindness of the bot community on Wikipedia. Yes, these things are part of the project's mission, but many in the community went above and beyond to help me understand their work and its significance. Thank you to everyone who participated in the study and helped me get the word out. Also, a special thanks to madman, Snowolf, Hellknowz, and the many IRC frequenters who helped by gathering data, interpreting policies, and answering little questions as they came up.

Thank you to my friends and colleagues who have sustained and encouraged me through the program at UO, including Brenna Wolf-Monteiro and the rest of my wonderful cohort, Katie Dettman, and Caitlin Mahy.

Most of all, my love and appreciation goes out to my family. My mother Barbara, sister Stacey, and brother-in-law Don supported my choice to go back to school, even though it took me about as far away from them as it could have. Thank you for your help and support through everything. And to my amazing wife Kimberly, who understands me better than anyone…you are my sunshine.

For my father, who would have enjoyed this more than anyone.

TABLE OF CONTENTS

Chapter Page

Chapter Page

Chapter Page

Chapter Page

Chapter Page

LIST OF FIGURES

Figure | Page

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

On April 18, 2012, congratulations from around the globe began appearing on the user Talk page of Wikipedia editor koavf,[1] who goes by Justin Knapp in the analog world. Fellow editor TonyTheTiger posted the first comment, a Special Barnstar[2] with the message: “Congratulations on becoming the first editor to 1 Million edits!!” Others well-wishers followed suit, and two days later, Wikipedia cofounder Jimmy Wales posted:

That same day, the online news site Mashable broke the story to the Internet community, claiming, “there’s not really a comparable accomplishment in the online realm” (Morris, 2012). Salutations continued to pour in with more barnstars, more exclamation points, and a thank you from the Wikimedia Foundation’s Executive Director Sue Gardner.[3]

One comment, however, was a bit unlike the rest. User Rcsmit posted: “Congratulations!!! But are all the (semi)-automated AWB-edits included?” AutoWikiBrowser (AWB) is a software tool used by many of Wikipedia’s power

[1] Many Wikipedia editors employ an all-lowercase spelling in their username.

[2] A barnstar is an informal award placed on a user’s Talk page to recognize good work or particular achievement on Wikipedia. Over 100 types of barnstars have been created since they were introduced in 2003, and since then they have “become ingrained in the Wikipedia culture” (Wikipedia, 2012q).

[3] The Wikimedia Foundation is the non-profit charitable organization that operates Wikipedia and a number of other wiki-based projects.

contributors that helps editors both identify pages in need of work and partially automates that work. Like the bevy of other automated and semi-automated tools available to the Wikipedia community, AWB was designed “to make tedious and repetitive tasks quicker and easier” (Wikipedia, 2012p).

To answer Kcsmit’s question, koavf’s one million edits do in fact include his semi-automated work, and a quick review of the 500 edits leading up to his milestone reveals that 259 (52%) of those edits were made using AWB and two other software tools, Twinkle and HotCat. And while the original question here was likely one of curiosity rather than criticism—indeed, one of Wikipedia’s founding principles is “Assume good faith”—it calls forth some significant issues around the nature of participation on the site, and the nature of online participation in general. Does it matter that this semi-automated work was not reported in the stories about koavf’s achievement? Is it merely assumed that software tools are used for this type of online work, or have these details become irrelevant? If so, why have they become irrelevant, and what are the implications of transparent technology? Should this type of work alter our perception of online participation and achievement? Would it even have been possible to reach such a level without automated assistance?

These questions have become important for understanding activity across the contemporary Internet. Unlike other media forms, digital media today is based around what Jenkins (2006) has called “participatory culture.” Users are creating the bulk of content on Facebook, YouTube, eBay, Twitter, and many of the world’s other most popular sites, while search engines like Google and Yahoo! are tapping into the Big Data[4] of users' browsing histories to offer targeted results and personalized advertising. Online shoppers can find a plethora of ratings and reviews on almost anything, and the blogosphere represents to some the promise of a reinvigorated public sphere.

Enabling much of this activity are technical enhancements that reduce barriers to participation. Programming skills are no longer needed by the average user to set up a WordPress blog or submit a review of a favorite restaurant to Yelp. Websites offer attractive page designs and integrated suggestions to make contributing as easy as a few clicks of the mouse. Consequently, although all of this activity is mediated through screens, keyboards, and touchpads, the technology itself seems to fade into the background. As the Internet becomes easier to use, the code, algorithms, and protocols—the very things that make up the Internet—move farther from view; *digital* media becomes *social* media as its technology becomes transparent.

Popularly dubbed the encyclopedia that "anyone can edit" with the purpose of collecting "the sum of all human knowledge," Wikipedia's image is that of an ultimate online reference tool which harnesses the wisdom of the crowd to author articles on everything from biochemistry to celebrity biographies. Unlike the corporate power plays in the information communication technology (ICT) sector that we see in the headlines quite regularly, with companies like Google, Apple, Verizon, and Comcast jockeying for control over both Internet access and Internet content, Wikipedia is an example of what the interactive Web could be if all information is free to access and free to alter. Internet scholars like Jonathan Zittrain (2008), in his popular Web 2.0 manifesto *The Future of*

[4] Lohr (2012), who documents its many current uses, defines Big Data as, "A meme and a marketing term, for sure, but also shorthand for advancing trends in technology that open the door to a new approach to understanding the world and making decisions."

the Internet and How to Stop It, laud the project for its open architecture and unrelenting community. Zittrain cites Wikipedia as a shining example of the "generative" Internet, a place where equal access and unvarying transparency define an environment much like the primordial circuits and code of the Net's childhood—open, decentralized, and organically driven.

But as the fanfare over koavf's achievement seems to indicate, transparency on Wikipedia is often subsumed by the dominance of its social sphere and article content. A broad and complex bureaucracy of policies, guidelines, permissions, and procedures have developed over Wikipedia's 11 years of existence despite the founding principles to "Be bold" and "Ignore all rules." Collective decisions are made using a consensus model, theoretically giving everyone an equal say, though in practice, experience and reputation play a major role in the community's dynamics. These social and bureaucratic factors also play a role in the development of content, from macro-level issues like what articles even deserve to be in the encyclopedia, to micro-level matters of punctuation and formatting.

All of these social and editorial aspects of Wikipedia are transparent in the fact that every on-wiki act is documented to the day, hour and minute, every change available for review and reversion. But what gets obscured, except to the most tech-savvy contributors, is the level of codependence between the social and technical structures of the project. For example, privileges and access levels are granted bureaucratically—a user is nominated for adminship[5] and promoted through the consensus process—but determined and enforced by the MediaWiki software that runs the site. Changes to those

[5] "Admin" is Wikipedia shorthand for Administrator, a particular type of user account with more advanced access and privileges than average registered users. "Adminship" is the state of being an Administrator.

privileges can be lobbied for by the community and rewritten in principle, but they are not actually materialized until they are recoded by MediaWiki developers and vetted for a number of technical considerations, a process that is itself both social and technical. Ultimately, then, to understand a system like Wikipedia, and digital media communities at large, we need to reinvigorate our discussions of cyberspace by paying precise attention to how social and technical structures constitute one another in order to form a sociotechnical system.

By considering the sociotechnical aspects of digital media systems, we unlock new ways to explore the collective intelligence (CI) that emerges from mass collaboration. Largely adapted from biological research on social insects, theories and concepts of CI are now being applied to cyberspace, with institutional collaboratives like the MIT Center for Collective Intelligence (2011a) investigating how "huge numbers of people all over the planet work together in new ways." Wikipedia, lovingly known by its community as a project that "works in practice, not in theory," provides fertile ground for exploring the mechanisms of CI precisely because contributors and their actions are transparently documented. By considering the sociotechnical implications for CI on Wikipedia, we can interrogate how individual intelligence is harnessed to create an emergent cyberculture that is simultaneously informational, technical, and social.

The purpose of this study, then, is to explore the codependencies of the social and technical structures that yield Wikipedia the website and Wikipedia the community. In doing so, the research investigates the implications of such a sociotechnical system for the maintenance of the project and the emergence of collective intelligence. To do this, we need to both look back at the history and development of Wikipedia, carefully examining the actors and forces that have shaped its trajectory, and closely examine the present social and technical infrastructures that sustain its dynamic activity. The following four research questions guide this inquiry:

RQ1: What is the historical context for the technical development of Wikipedia? What actors and forces have shaped its code, community, and infrastructure?

RQ2: In what ways is Wikipedia a sociotechnical system? What roles do the social and technical infrastructures of the site play in its collaboration and conflict? How does science and technology studies (STS) contribute to an understanding of the site's performance?

RQ3: In what ways is Wikipedia a system of collective intelligence? How is this frame applied and understood by the mass media? The research community? The Wikipedia community? How do other concepts of intelligence inform this discussion?

RQ4: What can a social network perspective reveal about the collaboration between human and non-human actors on Wikipedia? How could this collaboration be visualized? What implications emerge from this picture?

To address these questions, a largely qualitative multi-method research approach was employed, including document analysis, semi-structured interviewing, and social network analysis. Beyond its article content pages, Wikipedia is a vast archive of historical information documenting discussions and debates that have influenced its evolution. A plethora of documents were carefully selected and examined to explore how and why decisions were made, policies implemented, and technologies adopted on the site. Additionally, 45 interviews were conducted with members of Wikipedia's technical community to understand the relationships between social and technical aspects of the project and the motivations of programmers who contribute automated tools. Finally, social network measures and visualizations were used to interrogate notions of collaboration and make more transparent the centrality of technology to the content creation process. Collectively, these methods offered a triangulated approach for studying the sociotechnical aspects of Wikipedia.

Wiki 101

As the wiki has become both a concept central to understanding the current state of the Internet and a tool to facilitate interactivity in online networks, the term has been integrated into a vast number of names, expressions, and other concepts. Although each of the following terms is defined in more depth later in the project, a brief Wiki-101 here will help the reader begin to recognize differences in usage and understand how each is employed in this research project.

A *wiki* is an online technology used to collaboratively create and manage digital media content. It is accessed through a Web browser and generally available for anyone to work on. A wiki uses a simple, organized page design and archives all changes made to it, including who made the change, when the change was made, and what the change was. Consequently, a wiki is a simple database of content pages as well as metadata about that content. The *wiki concept* or *wiki philosophy* is used to describe the design features of this particular technology and the social media possibilities for mass collaboration on free and open user-generated content.

The largest and most well-known wiki is *Wikipedia*, a free Internet encyclopedia with over 21 million articles in 284 languages. Created in 2001 by Jimmy Wales and Larry Sanger, Wikipedia is currently a top-10 global website.[6] The project is written in many languages, and the name Wikipedia represents all language versions of the encyclopedia, which are hosted at the wikipedia.org domain, though each language

---

[6] According to Web metrics firm Alexa Internet Inc.'s widely-used rankings. Wikipedia has been in the top 10 since early 2007 and was #6 at the time of this writing.

version determines specific policies and protocols for that subdomain of the site. Wikipedia communities often overlap, with editors contributing to more than one language version, so although the present study primarily focuses on the English language version, “Wikipedia” will refer to the project as a whole, with specific versions identified as needed. “WP” will be used as shorthand for “Wikipedia”—for example, “English WP” will refer to the English language version.

A *Wikipedian* is a member of the Wikipedia community who writes, edits, or contributes to the project in some capacity. This includes the nearly 17 million named, registered accounts and an unknown (but relatively large) population of unregistered, anonymous contributors. Wikipedians are distinct from mere readers of the encyclopedia; one becomes a Wikipedian by actually adding something to the site, be it article content, discussion comments, or technical tools and scripts. On the site, *user* and *editor* are often used synonymously with *Wikipedian*, and all registered Wikipedians are given a *user page* to personalize and a *user Talk page* to interact with others in the community.

The *Wikimedia Foundation* (often shortened to “WMF” in this project) is the nonprofit charitable organization that funds and oversees a number of wiki-based websites, including Wikipedia. Established in 2003 by Jimmy Wales when Wikipedia was shifted from a dot-com to a dot-org domain, the WMF is now headquartered in San Francisco and employs 128 staff members and contractors, some of whom work remotely from around the globe. The *Wikimedia movement* is often used to encapsulate the WMF’s mission to bring free educational content to everyone around the world, but it also tangibly refers to the WMF’s websites, including Wikipedia, Wiktionary, Wikimedia Commons, and Wikinews, to name a few.

The WMF's sites are run using *MediaWiki*, a Web-based software application developed for Wikipedia. Originally produced by volunteers only, today MediaWiki is overseen by the WMF's paid technical staff, which collaborates with the larger volunteer community of developers on features and upgrades. Though most recognized for its use by Wikipedia and other sites in the Wikimedia movement, MediaWiki is a free, open-source program also used by private and corporate organizations to assist with collaboration and documentation of projects.

A (Very) Brief History of Wikipedia

Launched on January 15, 2001, Wikipedia began as a sister project of Jimmy Wales and Larry Sanger's online encyclopedia, Nupedia. Initially funded by Wales' successful Internet search company Bomis, Nupedia was to be an expert-written, freely distributed, advertising-supported online reference work. After less than a year, though, the project was struggling to produce content, so Wales and Editor-in-Chief Sanger decided to launch a wiki-based site to improve communication between contributors and experiment with online collaboration.

The wiki, a website software tool and design built around a philosophy of openness and simplicity, had been invented by computer engineer Ward Cunningham in 1995 to facilitate the documentation and workflow of his Portland, OR programming collective. By 2001, the wiki was used by a number of online projects, both private and public, as it featured a flexible structure that could scale with the size of a community. Wales installed the wiki package UseModWiki on a Bomis server in January of that year, and Sanger promoted the new project, Wikipedia, to various online mailing lists and bulletin boards. Wikipedia quickly gained an online buzz, as anyone, not just experts, could author the project's articles. By the end of the year, Wikipedia contained over 20,000 articles while Nupedia struggled to reach 100. Wales and Sanger soon turned their full attention to fostering their new success.

Wikipedia continued to grow in size and community over the following two years, with various language versions created around the globe. A volunteer developer community emerged around the project, and by 2003, the original UseModWiki software framework had been rewritten into a form more suitable for the encyclopedia: MediaWiki. Bomis suffered financially from the dot-com market crash, though, and Sanger, Wikipedia's only paid employee, left the project in 2002. After facing community resistance to the idea of selling advertising on the site, Wales decided to create the non-profit Wikimedia Foundation (WMF) in 2003 to oversee fundraising and development of the project. Wikipedia continued to flourish, reaching 1 million articles in 2004 and gaining traction in the new participatory Web environment.

With Wales as its chairman, the WMF operated with no paid employees until 2005, when longtime developer Brion Vibber was hired to oversee the technical development of the project. Relying heavily on corporate in-kind gifts and individual donations from its community, the WMF expanded Wikipedia's technical infrastructure from a few U.S. servers to multiple data centers around the world, while at the same time launching sister projects like Wikimedia Commons and Wiktionary. Still, the WMF prioritized maintaining a minimal organization whose role was to support the volunteer community and the Wikimedia mission to create and freely distribute educational content in the public domain. Over the next few years, a small staff was hired, including Executive Director Sue Gardner in 2007, and in 2008 the WMF relocated from St.

Petersburg, Florida to San Francisco, California in order to take advantage of strategic relationships with other Silicon Valley firms.

By the late 2000s, Wikipedia had emerged as a global phenomenon both online and in the offline world. Nearly 300 language versions had been created by millions of registered and anonymous Internet users, and Web traffic brought the site into the top-10 most visited on the Internet, with search engines often returning Wikipedia articles near the top of their results. Editors and contributors began holding face-to-face *meetups* in major cities to discuss their work, and two international conferences, Wikimania and WikiSym, have been held annually since 2005 to organize workshops, present academic research, and discuss a range of issues pertaining to Wikipedia and wikis.

More broadly, though, Wikipedia has become part of contemporary culture. Often criticized or lampooned in the popular media for its radical openness, the project calls into question established notions about truth, objectivity, and authority. Schools and libraries are developing curriculum around the site that both teaches critical media literacy and fosters reading and writing skills. Corporations and politicians are turning to public relations specialists to make sure their articles are as "clean" as possible. Net neutrality advocates are rallying around Wikipedia as the proof of their cause. Indeed, as Wikipedia turns the page on its first decade, it stands alone as the most prominent, most recognized symbol of openness and free culture in an increasingly closed, commercialized online landscape.

Significance of the Study

As Wu (2010) documents in *The Master Switch*, virtually all modern communication technologies—from the telephone to cable television—have gone through a distinctive pattern beginning with open, sometimes chaotic innovation and ending with monopolistic or oligopolistic control by the owners of the technical infrastructure. He sees this pattern currently playing out with the Internet, and along with several other authors, is worried by the prospect of the Internet becoming a closed system that operates almost purely according to market logic. Whereas these authors focus on the traditional centralizing and conglomerating tendencies of communication industries, others like Lessig (2006), Galloway (2004), and Terranova (2004) look to the technical aspects of online communication—code and protocol—to describe how control on the Internet is maintained in spite of its decentralized and distributed structures. This approach is based on "the recognition of a newly salient regulator … the software and hardware that make cyberspace what it is also regulate cyberspace as it is," and its distinctiveness lies in the fact that it considers both humans and the advanced digital technologies they create as powerful actors in the current transformation of a "cyberspace of anarchy to a cyberspace of control" (Lessig, 2006, p. 5).

The non-profit Wikipedia is popularly cited as an online anomaly, a top global website with an open and transparent organization that is largely free from social, economic, and technical pressures. However, experience on the site and recent research into its content and community tell us this is not the case. The site of both collaboration and conflict, Wikipedia is a complex network of the social and the technical, with human and non-human actors both contributing to a project that, taking a page from old *Encyclopædia Britannica* ads, the site's founder has dubbed "the sum of all human knowledge." Critical of the normative Web 2.0 doctrine, this dissertation uses Wikipedia as a case study to explore the broader questions around digital media platforms, including

"How do social and technical forces shape behavior and content in the online communication space? What issues of agency and control arise? And does something greater than the whole—collective intelligence—emerge from the resulting network of actors?" These questions are important for locating the significance and development of digital media in the tradition of mass communication technologies and for addressing concerns that the Internet will not ultimately maintain its open architecture and democratic promise. Though this project is a case study, the issues addressed in it are common to many digital media systems. Thus, the insights and conclusions presented here can serve to inform broader studies of the online environment.

This research also provides the first extensive and in-depth study of software robot (bot) programmers on Wikipedia. With trends suggesting the project's editor base may be in decline (Ortega, 2009) and estimates indicating that bots contribute nearly 22.5% of all edits to the site (Zachte, 2011), bot programmers are in a unique position to influence the future direction of the site. Bots perform work that is largely undesirable to human editors, sometimes even unimaginable in terms of scope and duration. By understanding how and why bot programmers maintain a system of "zombie labor"[7] behind the scenes to support the more nuanced work of human editors, we can more fully understand and appreciate the dynamic sociotechnical system of collaboration that drives the site towards the sum of all human knowledge. Inspired and informed by earlier quantitative and trace ethnography work on Wikipedia bots by Geiger (2010; 2011), this dissertation research gives voice to an important sub-community of Wikipedians and

[7] "Zombie" is a term employed by User Cyde in reference to Cydebot, one of the most active bots on Wikipedia. He uses "zombie labor" to more generally describe the overall work of bots on the project.

brings their issues, concerns, and successes out of techie-only chat rooms and into the broader Internet discourse.

Overview

This study draws on a number of theories and interdisciplinary literatures to understand the sociotechnical nature of Wikipedia and its implications for collective intelligence on the site. Chapter II discusses theoretical frameworks, which include new/digital media theory, information theory, network theory, science and technology studies, and the political economy of communication, that are essential to an understanding of the wiki phenomenon. In addition, major studies and dominant themes in wiki and Wikipedia research are presented to describe existing approaches and provide a basis for this project's methods. Chapter III documents and justifies the three methodologies employed in this dissertation—document analysis, interviewing, and social network analysis—while also commenting on the new and evolving nature of Internet research.

Chapters IV through VI present the major findings of the dissertation. Chapter IV presents background context for the emergence of Wikipedia, including the impetus for encyclopedic projects, the need for information in contemporary society, and the rise of the Internet. The chapter then uses the heuristic of a wiki page to present a technical history of Wikipedia's development, emphasizing the technological, social, political economic, and ideological actors and forces that have influenced its growth.

Chapter V argues that Wikipedia is best understood as a sociotechnical system of human and technological agents. The chapter begins with a functional analysis of the social and technical infrastructures on the site and provides an explanation and description of relevant groups, procedures, and technologies that contribute to its day-to-day functioning. A sociotechnical analysis of the entire system is then offered to detail the durable yet dynamic ways in which these groups work together, shape one another, and hold each other in place. Throughout this chapter and the next, bots are discussed as specific examples of the sociotechnical nature of Wikipedia. An exploratory network analysis concludes the chapter, offering a glimpse into the centrality of automated tools to the maintenance and improvement of the encyclopedia's content.

The specific implications of a sociotechnical perspective of Wikipedia for the possibilities of emergent collective intelligence are then explored in Chapter VI. After reviewing how theories of intelligence in humans and computers inform notions of CI, the chapter discusses how this concept is understood and applied by various publics associated with the project. Next, specific mechanisms and models of CI are held up to Wikipedia, revealing that the system largely does exhibit qualities of stigmergy, distributed cognition, and emergence. The chapter concludes by commenting on the unique possibilities for cyberculture at the convergence of digital sociotechnical systems and mass collaboration.

Finally, Chapter VII summarizes the study's findings and major assertions, acknowledges important limitations, and suggests new and future directions for understanding Wikipedia and collaborative digital media systems as sites of sociotechnical work.

CHAPTER II

THEORETICAL FOUNDATIONS AND LITERATURE REVIEW

In order to explore theories of social and technical organization on a recent and still emerging digital communication system like Wikipedia, a review of foundational literature and relevant historical developments is necessary to ground subsequent arguments and ideas advanced in the present research. This chapter provides such a base by examining established theories and research from diverse disciplines that inform a sociotechnical analysis of collaboration and intelligence on Wikipedia. To being, digital media theory is considered, from its roots in early information theory to its applications in modern mass media. Principles of network theory relevant to this study are then reviewed, followed by major concepts and perspectives from the interdisciplinary field of science and technology studies (STS). The political economy of communication is then discussed, focusing in on analyses of online communication systems and digital labor. Finally, a brief survey of research on online communication, including major work on Wikipedia itself, is offered to frame the significance of this study's research questions and approach.

New/Digital Media Theory

Foundations of Information Theory

Before approaching contemporary definitions and applications of digital media theory, it is important to understand where many of the theoretical and technological concepts that enable digital media originated. At the heart of this understanding is an interrogation of *information* from both technical and philosophical standpoints. Day (2001) argues that the term was largely transformed during the twentieth century, its connotation changing from the notion of a “process” (imparting knowledge) to the notion of a “thing” (a piece of knowledge). As it pertains to digital media, this transformation was closely tied to the development of information theory and cybernetics in the U.S. and U.K. during the post-World War II period.

The Mathematical Theory of Communication. Shannon’s (1948) landmark paper “The Mathematical Theory of Communication” was a key starting point for digitized media and communication, as Shannon developed a theoretical system by which Boolean algebra (X and Y, X or Y, X not Y, etc.) could be conducted via electrical circuits to establish a chain of communication. The pattern of communication of interest in his study was linear, and his main concerns, influenced by his work with the U.S. government during World War II, were for the costs of transmission and the ability to overcome disturbances (or “noise”) in the chain (Figure 1). Building from the work of Ralph Hartley and Alan Turing on binary information, though, Shannon’s more pivotal contribution in “The Mathematical Theory of Communication” was the conceptual system he introduced that allowed for the physical embodiment of logic (and by extension, thought) in the digitized form of *bits* of 0s and 1s. This purely logical electronic translation also served to divorce meaning from information, as the model was only interested in the reliability of message transmission, ignoring “the meaning of the signals … how they are understood by the receiver, or the intention behind the transmission” (Mattelart & Mattelart, 1998). For this theorizing on the distinction between the content of communication and its media of production, transmission, and reception, published in the wake of World War II while working at the famous Bell Laboratories, Shannon has become known as the “father of information theory” (Gleick,

2011) and has been a key influence on the development of ARPANET, packet switching, and what we now know as the Internet (Fisher, 2010; Galloway & Thacker, 2007).

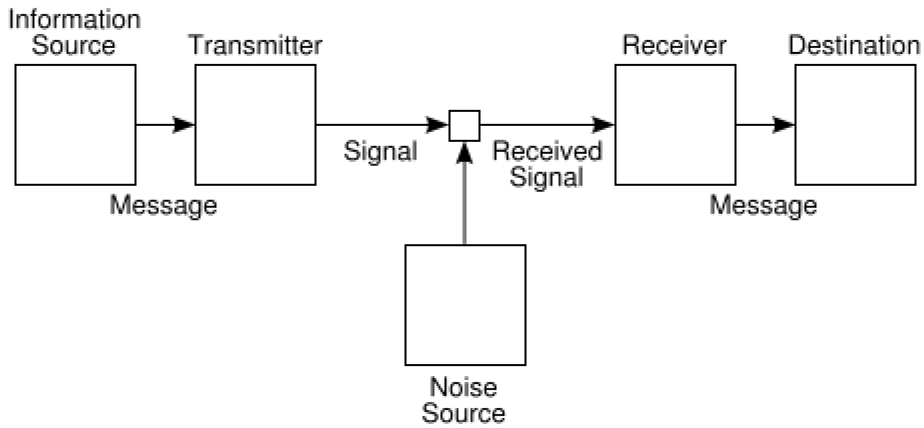

Figure 1. Shannon's diagram of a general communication system. Image from User:Wanderingstan and User:Stannered, Wikimedia Commons

A year after Shannon's treatise appeared in *Bell System Technical Journal*, a co-authored book with Warren Weaver (1949) on the theory was published. Whereas Shannon's interest lied more in the technical and engineering aspect of the model, Weaver was interested in expanding these ideas on message transmission into a more general theory of communication, and thus, he was interested in the semantic problems of meaning and interpretation that the model filtered out (Day, 2001). Weaver suggested that these subjective aspects of the message are largely subsumed in the technical problems of distribution, and that "an informational reading of sensory, emotive, or cognitive affect reduces all affective events to being effective events (thus requiring an intentional or causal subject-object relationship and introducing issues of probability, measurement, noise and delay, and feedback)" (Day, 2001, p. 41). What emerged from this conceptual model of communication, and what Weaver and cyberneticist Norbert

Wiener struggled with, was a deterministic system that was both descriptive and prescriptive.

Cybernetics. Working in the same postwar environment and scientific/engineering community as Shannon and Weaver, Wiener (1961) published *Cybernetics, or Control and Communication in the Animal and the Machine*. As he describes in the book, cybernetics (from the Greek, meaning "steersman") largely grew out of concerns for command and control in engineering during the World War II, but its insistence on constant information and communication exchange provided a framework for studying regulatory systems, both physical and social. Similar to Weaver's information theory, "the logic of cybernetics is that of systems engineering, which means that language and affects are viewed in terms of systems, quantitative values, message transmission and effects, and management and control" (Day, 2001, p. 49). Cybernetics assumes a hostile, entropic environment where constant feedback from that environment is necessary to maintain control and facilitate purposeful communication. Wiener also wrestled with the duality of a cybernetic system being both a maintainer of order and a determinant of social spaces (Day, 2001), but modern applications of cybernetics are often more functional than philosophical, and its attention to systems of control are mainly used in the fields of computing and engineering, biology, and management.

Day (2001) writes that information theory and cybernetics are both theoretically and pragmatically useful foundations for communication because they strive for "the clear transmission … of reason into practice, thought into the world" (p. 54). These ideas would facilitate the development of computer hardware and software in the latter half of the twentieth century and inform the work of many contemporary digital media theorists.

The Development of Computer Hardware and Software

Spurred by the work and ideas of Shannon, Weaver, Wiener and other World War II-era mathematicians and engineers, the Information Revolution paradigm emerged as a dominant technical and economic (and later cultural) ideology for the remainder of the twentieth century and into the twenty-first. This development will be explored in more depth in Chapter IV, but it is important to briefly discuss here how ideas from information theory enabled the rapid growth of computer technology, as this technology in turn enables our contemporary digital communications.

Hardware. Over a century before the development of information theory, British mathematician and engineer Charles Babbage imagined calculating machines—*computers*—that would provide the basic architecture for the modern digital computer. His two primary visions, the difference engine and the analytic engine, were both mechanical machines that used gears, wheels, and bearings to make calculations based on mathematical logic, and neither machine was fully constructed before Babbage died in 1871[8] (Hillman & Carr, 2011; Slater, 1987). However, Babbage is deemed the "father of the computer" for his theoretical design of the analytic engine (Halacy, 1970; Markoff, 2011), which included an arithmetic-logical unit, a memory unit, a control unit, and an input/output unit, all precursors for elements of a modern computer (CPU, RAM and hard drives, keyboards, monitors and printers). The analytical engine has come to be known as the first Turing-complete design for a computer (Graham-Cumming, 2010).

The conceptual work of Alan Turing in the early twentieth century, as well as his contemporary Jon von Neumann, provided additional foundations for the development of

[8] In fact, the analytic machine was never more than a detailed set of blueprints and drawings during Babbage's lifetime.

digital computing. Though largely ignorant of Babbage's invention, Turing invented a similar conceptualization for a computing machine, but one where the universal adaptability of the machine is the theoretical focus (Markoff, 2011). His Universal Turing Machine is a conceptual, symbolic machine intended to manipulate other symbolic machines; it contains a tape of written symbols, a head to read the tape, a table of rules and instructions, and a register to store the state of the machine (Kirschenbaum, 2008; Turing, 1936). The "computer" does not need to be a machine; it could also be a person mechanically following the rules of the system. Ultimately, the Universal Turing Machine demonstrated "the important characteristic of a computer is that it can mimic any machine, any piece of hardware, provided that the functionality of that hardware can be broken down into logical processes" (Galloway, 2004). Working both independently and with an awareness of Turing's ideas, mathematician Jon von Neumann proposed a streamlined stored-program design that stores program data and instructions in the same internal computer memory; the first electronic computers in the late 1940s and 1950s employed this "Von Neumann Architecture" (Rojas & Hashagen, 2000; von Neumann, 1945).

Computing hardware exploits both the theoretical designs of Babbage and Turing and the logical manipulations of Shannon's information theory; both the larger, machine-level ideas of the former two and the micro, data-level ideas of the latter are necessary for a digital system. Early computation machines implemented many of these ideas with existing technology, including the use of punched cards in lieu of internal memory and wire switching or vacuum-tube systems (borrowed from radio and television technology) for logical operations, though none of these were fully electronic systems. It wasn't until the invention of the transistor in 1947, followed by the subsequent development of silicon chips, integrated circuits, and microprocessors over the next thirty years, that computers could quickly and efficiently manipulate electronic pulses in the way prescribed by Shannon's vision, utilizing internal memory as von Nuemann imagined.

Software. Contrary to the conception that computer software is a fundamentally separate entity than computer hardware, Ceruzzi (1998) argues:

> A computer system is like an onion, with many distinct layers of software over a hardware core. Even at the center—the level of the central processor—there is no clear distinction: computer chips carrying "microcode" direct other chips to perform the processor's most basic functions. (p. 80)

Thus, software was mutually developed with hardware as computing technology progressed in complexity.

At its simplest, software is a set of instructions to direct a computer (Ceruzzi, 1998). Early computers without internal memory used punched cards to feed instructions to a machine; the presence or absence of a "punched" hole in the card represented one bit of data (though later card designs used punched cards in more advanced ways). Rolls of tape were also sometimes used both to input software instructions and record outputted data. These external media had numerous drawbacks, however, including the costs of the materials, deterioration and data loss, and the physical labor needed for inputting instructions (Campbell-Kelly, 2003). The development of internal circuitry and memory eliminated these problems and enabled the development of more sophisticated software, including operating systems that manipulate basic machine language and function, and programming languages that use precise syntax and advanced algorithms to maximize resources (Campbell-Kelly, 2003; Ceruzzi, 1998).

An important feature in software's role in computing, and one examined by new/digital media theorists like Manovich (2002) and Galloway (2004), is its ability to manipulate information without concern for the underlying representation—the "what it means"—of that information. Known as functional abstraction:

> the details of the algorithms to accomplish the function are not visible to the consumer of the function. The consumer of the function need only know the correct calling convention and have trust in the accuracy of the functional results (Burback, 1998).

Computers operate on a hierarchy of abstractions, built up from the basic manipulation of individual bits at the bottom to the top-level programs that users interact with. Indeed, contemporary computers are so complex that even programmers rarely work on lower-level abstractions directly. Conceptually, functional abstraction fulfills a crucial operation of Shannon's (1948) mathematical theory of communication: it divorces meaning from information. Ultimately, computer hardware and software work together to offer systems of physically embodied logic that are fundamentally uninterested in the greater meaning behind that logic; the implications of this are dynamic and debated in many of the fields affected by computer technology, including communication and media, while the similar conceptual construct of the *black box* is key to the field of science and technology studies (STS).

Contemporary New Media Theory

Although *new media* as we understand the term today has only been in the communication lexicon for a few decades now, we see that some of the fundamental elements that inform new/digital media were envisioned much farther back in fields outside of mass communication. Our current understandings of new media as it pertains to our communication experiences is still somewhat nebulous, however, as few would argue the scientific simplicity of Shannon's model and what it enabled for digital communication is anything but the beginning of the story. Indeed, the difficulty in defining new media is itself a topic of debate, as Lievrouw and Livingstone (2002) point out:

> The field needs a definition that is abstract enough to accommodate the range of systems, contents, issues and settings that researchers consider essential, yet not so broad that new media cannot be distinguished from other established areas within communication research and other disciplines. (p. 5)

No perfect definition exists, and new media theory is ultimately both an effort to understand the experiences and artifacts within its boundaries, and an ontological exploration of where those boundaries are exactly.

Still, after surveying the dominant discussions of new media, Lievrouw and Livingstone (2002) offer the framework for a definition: "By new media we mean information and communication technologies and their associated social contexts," consisting of the artifacts or devices themselves, activities or practices of engaging with these items, and the social arrangements that form around them (p. 7).

New Media Form. We can begin talking about new media and its forms as they relate to other types of traditional media, a method utilized by Bolter and Grusin (1996; 2000) and Manovich (2001), among others. Bolter and Grusin (2000) build a strong historical case to argue that "remediation is a defining characteristic of the new digital media" (p. 5). Remediation is the refashioning of previous media in a new context, and it is not new to digital media; the authors present a number of examples, including photography remediating painting and film remediating stage plays, to argue a historical pattern in the development of media. For Bolter and Grusin (2000), new digital media merely represents a media form that enables an extreme ability for remediation based on the “twin logics of immediacy and hypermediacy” (p. 5). Digital media offers an intimacy that can psychologically or physically immerse the audience in a situation or experience, and yet to do this, it must rely on a confluence of media conventions and functions, both technical and cultural. New media, then, is old media experienced in a fundamentally new way.

Manovich (2001) has gone further in defining the formal aspects of new media objects and is one of the key figures in new/digital media theory. In *The Language of New Media* (2001), he outlines the “Principles of New Media” by presenting five key differences between old media and new media, as well as a number of popularly held notions of new media that he finds false. Of these key traits, two prove most significant for Manovich. Because new media objects operate with *variability* (the ability to appear in numerous forms), some of the authority over the objects is passed along to the user. The implications of this characteristic are twofold. At the phase of production, issues of labor arise as the development of new media objects and media content can largely fall on users/consumers, enabling media owners to maximize profits from surplus labor, whether it be traditional, affective, cognitive, or immaterial labor[9] (Jenkins, 2006; Terranova, 2004).

Of more interest to Manovich, though, are the possibilities of use (we might say “consumption,” though a flavor of consumption much different than with old media) that variability allows, and he highlights menu interactivity and hypermedia as emblematic instances of this principle. Variability allows users to play an active role in the order that media elements are accessed, to create and/or manipulate elements, and to interact with networks of information to create personal experiences. On a much larger level,

[9] This point will be discussed in further detail later in this chapter.

Manovich (2001) uses the variability principle to exemplify how “changes in media technologies are correlated with social change”:

> If the logic of old media corresponded to the logic of industrial mass society, the logic of new media fits the logic of the postindustrial society, which values individuality over conformity. In industrial mass society everyone was supposed to enjoy the same goods—and to share the same beliefs. This was also the logic of media technology. A media object was assembled in a media factory (such as a Hollywood studio). Millions of identical copies were produced from a master and distributed to all the citizens. Broadcasting, cinema, and print media all followed this logic.
>
> In a postindustrial society, every citizen can construct her own custom lifestyle and “select” her ideology from a large (but not infinite) number of choices. Rather than pushing the same objects/information to a mass audience, marketing now tries to target each individual separately. The logic of new media technology reflects this new social logic. (p. 41-42)

Variability is a mechanism by which these logics are realized in digital media.

The second important principle of new media described by Manovich (2001) is *cultural transcoding*, “the most substantial consequence of the computerization of media,” as it acknowledges a dual nature in new media objects (p. 45). These objects exist both as cultural products (i.e. the content of a photograph as we see it with our eyes) and as digitized products (i.e. the content of the photograph in bits, pixels and hard drive locations). Manovich stresses that the dual nature of computerized media calls for a new theoretical perspective that recognizes our media is now “rendered” at the same time in both the cultural and technical realms; a great deal of media is created, stored, distributed, and played via computer technology, and this affects both the media product itself and how we see and perceive the media product. Here Manovich (2001) theoretically brings back together the notions of form and content that Shannon’s (1948) information theory served to separate, though they are not reunited fully, but rather dealt with as a consequence of new media’s “programmability … the most fundamental quality of new media that has no historical precedent" (p. 47). Galloway (2004) and Lessig (2006) elaborate on this idea, arguing that Internet protocol and programming code, respectively, are key factors that influence how content is produced and consumed.

Manovich (2001) is careful in his analysis not to forget aspects of traditional media that are present in new media, or to inflate new media based on aspects or attributes that are not indeed new. He describes and debunks certain myths about new media that seem to exaggerate its importance; among these myths are that new media technologies do not suffer degradation in the replication or distribution process (when in fact, files are often compressed and lose data), or that new media technologies are the most precise technologies (when in fact, certain cameras can actually capture more detail optically than digitally). The most important myth Manovich addresses, however, is that of the interactive nature of new media. Touted by Jenkins (2006), Benkler (2006), Levy (2001) and others as one of the prized aspects of new media, Manovich reminds us that interactivity is not merely a technical feature (although interactivity was possible in technical terms with other media, from letters to the editor to call-in shows), but also a psychological feature. The author reminds us that "the psychological processes of filling-in, hypothesis formation, recall, and identification, which are required for us to comprehend any text or image," have always accompanied the consumption of media objects, both old and new (Manovich, 2001, p. 57); it is not the mere interactivity of new media that is unique, but the form of interactivity (based on the variability principle) and the control of the experience that distinguishes digital media.

Online Engagement. As Manovich makes clear in his analysis, form, function, and consumption of new/digital media are far from mutually exclusive areas of theory and research. Neither does the form of new media fully determine how users engage with it, however, with the Arab Spring uprisings a prime example of unanticipated technology use.[10] Still, the new forms of interactivity and engagement enabled by new media have played a major role in theory building.

Further developing the ideas on remediation put forth by Bolter and Grusin (1996; 2000), Deuze (2006) argues that participation is a principle component of the digital experience, and even more specifically, that the proliferation of online access and usage has given media consumers the ability to tailor their own experiences using Levi-Strauss's concept of *bricolage*. Focusing on the World Wide Web, he points out: "Bricolage is evident in the ways in which we click, publish, and link our way online" (Deuze, 2006, p. 70). For bloggers and online journalists, the primary object of Deuze's study, the implications of participation and bricolage are varied; it can be argued that personalized news and information can create a personalized "truth," a dubious proposition in the field of journalism, but Deuze (2006) sees participation through bricolage as a socially critical component of new media—a mitigating control over the largely uncontrolled media and information environment thrown at us each day. In either case, digital media has spurred "the emergence of new types of citizenship, participation, activism, dialogue, and interactive communication" (Deuze, 2001, p. 72).

Bruns (2007; 2008) has elaborated further on the significance of participation in new media. Examining diverse realms of the online world, including the blogosphere, Wikipedia, and Second Life, Bruns (2007; 2008) characterizes engagement with new

---

[10] Journalists and researchers are now assessing the role that ICTs like Facebook and Twitter have played (and are continuing to play) in the organized protests and democratic revolutions in Northern Africa and Western Asia (see Essam, 2012; Rosen, 2011; Technology Review, 2012; etc.). It is clear, however, that such political uses were not originally intended for these social networks.

media objects as *produsage*, a portmanteau of "producer" and "usage" that "encapsulates the paradigm shift towards user-led forms of collaborative content creation which are proving to have an increasing impact on media, economy, law, social practices, and democracy itself" (Bruns, 2007). His creation of a new term for this type of online engagement is an effort to further distinguish the logic of industrial mass society and its methods of production from the logic of postindustrial society and its production, as Manovich highlighted earlier.

Virtual Communities. Flew (2002) writes that the study of new media needs to stress "how the mediation of communications through technological forms renders communications a form of social practice" (p. 10). Indeed, the social aspects of new/digital media technologies have come to the forefront in the Interactive Web (or Web 2.0) era, and as McQuail (2005) argues, "new theory is only likely to be required if there is a fundamental change in the forms of social organization of media technologies, in the social relations that are promoted" (p. 136). Jankowski (2002) sums up both the interactive engagement and social community aspects of new media in summarizing the debates found in the first issue of the journal *New Media & Society*: "New media are, to a large degree, socially constructed phenomena and often deviate substantially from the designer's original intent" (p. 35).

An early and notable proponent of virtual communities and their implications for mass communication and society, Rheingold (1993) was himself a member of an early online community where people could "do just about everything people do in real life, but we leave our bodies behind" (p. 3). His thesis on why a growing number of people have migrated to online social spaces revolves around the loss of offline community;

building from Putnam's (1995) sociological work on the decline of civic engagement and social participation, Rheingold (2000) argues that informal public spaces for discussion and camaraderie are disappearing from our real lives, but the ease of access to both people and information offered by the Internet can play a major role in reinvigorating community values. Writing at a time when "the Net [was] still out of control in fundamental ways," the author worries these online spaces will soon be corralled, as "big power and big money always [find] ways to control new communications media. … [They] seize it, censor it, meter it, and sell it back to us" (Rheingold, 2000, p. 6).

Writing and researching over a decade later, Jenkins (2006) explores ideas of virtual community in the contemporary, commercialized Internet era that Rheingold feared. He claims we live in a "convergence culture, where old and new media collide, where grassroots and corporate media intersect, where the power of the media producer and the power of the media consumer interact in unpredictable ways" (Jenkins, 2006, p. 2). Rather than falling victim to corporate media interests, Jenkins argues that virtual communities use new media to both consume and resist dominant ideologies, negotiating between these two poles as they see fit. As new media is highly participative, communities leverage their collective intelligence via technology to make media a highly socialized experience, one that Jenkins (2006) links back to oral traditions where audience participation played a key role in the creation of stories.

As we've seen, the roots of new/digital media stretch much farther back than the explosion of the Internet and communication technologies over the past fifteen years. Ideas and principles from information theory, cybernetics, and the reciprocal development of computer hardware and software are important for understanding current strains of new media theory, which attempt to identify and explain media forms, user participation, and online community. Each of these areas plays a significant role in understanding both the wiki and Wikipedia.

Social Networks

The emergence of digital/new media has sparked an outbreak of research and literature on networks and their implications for mass communication (Barabasi, 2002; Benkler, 2006; Castells, 1996, 2004; Christakis & Fowler, 2009; van Dijk, 1999; Watts, 2003). Though this work sometimes considers the hardware infrastructure that enables the network of networks known as the Internet, more often it seeks to understand the social formations and behaviors of actors in a network. Recent websites like Facebook and MySpace have popularized the notion of *social networks*, but the theories and understandings of such social structures have a much longer history in the offline world. These established ideas can help us make sense of online experiences and the networked collaboration that drives Wikipedia.

Network Theory and Basic Concepts

Network theory is formally based in the fields of mathematics and graph theory, where it is used to understand the representation of relationships between discrete objects (Newman, 2010). Famous mathematical problems like the "Traveling Salesman Problem" utilize network theory to understand real-world analogies—in this case, the shortest possible route for a salesman to visit every city along his route exactly once (Applegate et al., 2006). Network theory's focus on relationships has proven useful in many disciplines, and network structures and measures are used in research from neurobiology to ecology, computer science to sociology. Even philosophy and the humanities have used network principles to understand the relationships between words, arguments, and ideas.

Networks are composed of nodes (or *actors*) and relations (or *ties*), and network analysis is keenly interested in the relationships (or *connections*) between nodes. Traditional data analysis generally emphasizes actors and their attributes rather than their relationships, with Hanneman and Riddle (2005) pointing out "the difference in emphasis is consequential for the choices that the researcher must make in deciding on research design, in conducting sampling, developing measurement, and handling the resulting data." Networks have a shape (or topography) based on connections that are present or absent, and this topography can be measured mathematically and represented visually to reveal information about the network. Measures of density, connectivity, and reachability can be calculated for the network as a whole, and measures of centrality, closeness, and betweenness can be calculated for each individual node. The fact that networks lend themselves to visual and spatial representation, though, gives the researcher an additional technique to analyze and explore nodes and relations. Numerous software tools are available for network analysis "to develop insights based on characteristics of the data, … to conduct preliminary tests of a priori ideas, to explore the fit of models to data and, using animation, to examine dynamic processes" (Freeman, 2005, p. 268).

Social Networks and Pioneering Research

Social networks apply the concepts of network theory to more literal actors, usually humans. Christakis and Fowler (2009) describe how a social network is fundamentally different from a group of people:

> A group can be defined by an attribute (for example, women, Democrats, lawyers, long-distance runners) or as a specific collection of individuals to whom we can

> literally point ("those people, right over there, waiting to get into the concert"). A social network is altogether different. While a network, like a group, is a collection of people, it includes something more: a specific set of connections between people in the group. These ties, and the particular pattern of these ties, are often more important than the individual people themselves. They allow groups to do things that a disconnected collection of individuals cannot. The ties explain why the whole is greater than the sum of its parts. And the specific pattern of the ties is crucial to understanding how networks function. (p. 9)

Social networks are continuously evolving and self-constituting, which makes them distinct from the networks of mathematics and graph theory (Watts, 2003). In addition, nearly a century of research on social networks has revealed some key principles of this evolving self-constitution: we shape our network, our network shapes us, our friends affect us, and our friends', friends', friends affect us (Christakis & Fowler, 2009).

What is now understood as social network analysis began in the 1930s with the sociograms of Moreno (1934) and studies by Davis and others of social circles in American communities (Scott, 1991; Wasserman, Scott, & Carrington, 2005). Travers and Milgrim (1969) tested the bounds of an individual's social network on a much larger scale with their famous "Small World" experiment; they tracked how many personal acquaintances were required to get a letter from a person in Nebraska to a person in Boston, arriving at a final average of six. Though only sixty-four letters reached their ultimate target, quite a small number to generalize significant results from, "Six Degrees of Separation" became a rule of thumb for the closeness of any one person to another, and the experiment's results were further replicated and confirmed (Watts, 1999). Dodds, Muhamad, and Watts (2003) updated the Six Degrees test on a global scale using email, finding again that the average number of steps to get a message from the random starting point to the targeted receiver was six. Christakis & Fowler (2009) later contend that three is actually a more significant number; although there are six degrees of acquaintance between any two people, influence—defined as attitudes, feelings, and behaviors—only spreads out three degrees from the source.

Granovetter's (1973) work on the strength of network ties presents another important trait for understanding network interaction. Arguing "a fundamental weakness of current sociological theory is that it does not relate micro-level interactions to macro-level patterns in any convincing way," Granovetter (1973) explores the importance of weak ties in connecting large networks. He concludes that, paradoxically, weak ties often associated with alienation from local relations are "indispensable" to maintaining cohesion in a larger community, and ultimately, "the personal experience of individuals is closely bound up with larger-scale aspects of social structure, well beyond the purview or control of particular individuals" (Granovetter, 1973, p. 1377-1378). This condition of social collectives is quite similar to the concept of *emergence* in the field of collective intelligence, which is discussed generally and in the context of Wikipedia in Chapter VI.

The Networked Society

It is important to note the rise of the *network discourse* that has accompanied the latest stage of the information revolution, and its relation to network theory. Castells (2004) describes a "network society … whose social structure is made of networks powered by microelectronics-based information and communication technologies" (p. 3); van Dijk (1999) contends that "the basic elements of the network society are not so much networks themselves but individuals, households, groups and organizations linked by these networks" (p. 24); and Terranova (2004) writes of "a cultural formation, a network culture, that seems to be characterized by an unprecedented abundance of informational output and by an acceleration of informational dynamics" (p. 1). These authors and others

(Barabasi, 2002; Benkler, 2006; etc.) use concepts and explanations from network theory to elucidate what they view as significant trends and changes to the way society is ordered both socially and economically, but the rhetoric often moves well beyond the scientific analysis that network theory is suited for. The term *network* has become politically charged, distinct from the original mathematical usage of the term. The materiality of networks plays a role as well; network discourse flows from the integration of the Internet and ICTs into daily life, and the material connectedness of digital/new media serves as a predominant metaphor for the psychological connectedness that emerges. The promise of digital/new media systems like Wikipedia lies in its networked nature, both literally and figuratively, and as the following section will illuminate, science and technology studies has offered new and important perspectives for understanding this sociotechnical nature.

Science and Technology Studies

Writing in 1992, Bijker and Law argue:

> the academic time is right for work on the sociotechnical. … We are witnessing the birth of a new capacity to understand, in a matter-of-fact way, how it is that people and machines work together, how they shape one another, how they hold one another in place. (p. 306)

Indeed, the last twenty-five years has seen the emergence of science and technology studies (STS) as a theoretical framework for investigating the convergence of science, technology, and society. Besides its applications to the traditional natural sciences, this framework is also used in the fields of organizational analysis, labor studies, engineering, and most pertinent to this study, digital media and Internet studies.

Social science and humanities research certainly considered science and technology prior to the emergence of STS, but most of this historical, philosophical, and sociological work divorced science and technology from its social contexts:

> Philosophers studied the logic of the scientific method, Whig historians documented the natural evolution of ideas and technological artifacts, while sociologists gazed at the institutional structure of science and its pattern of communication and reward. (Bowden, 1995, p. 70)

Three major breaks from these traditional approaches signaled the beginning of STS as a field. First, historical accounts of science began to be contextualized by the real-world problems they addressed, and the social impacts of the solutions they offered, including ramifications for public policy, were emphasized. Second, the content of scientific knowledge was opened up to sociological scrutiny that challenged the epistemological authority of traditional science as a reflection of nature. Finally, a “turn to technology” connected critical inquiries into science with those of applied technology, with academic programs and journals beginning to institutionalize STS perspectives (Bowden, 1995, 70-72). Today, university departments or programs specifically defined by the questions of STS thrive at many science and technology focused institutions, including Cornell University, Rensselaer Polytechnic Institute, Massachusetts Institute of Technology, Virginia Tech University, York University, and University of London, to name a few.

STS is grounded in the assumption that “science and technology are thoroughly social activities,” where different social groups, behaviors, preferences, ideas, and contexts influence both processes and outcomes (Sismondo, 2004, p. 10). STS opens its analysis to more than merely scientists and the objects and artifacts they work with; it recognizes that scientists are “people, not minor deities” who work within a much larger network of actors whose economic, political, and cultural interests and biases affect scientific and technological progress (Fuller, 2006, p. 2). More philosophically, STS questions the positivist view that science is a natural process that reveals natural truths inherent in the world. Instead, the perspective argues: "The sources of knowledge and artifacts are complex and various: there is no scientific method to translate nature into knowledge, and no technological method to translate knowledge into artifacts" (Sismondo, 2004, p. 10). Ideas on the social construction of knowledge, then, play a significant role in making sense of scientific accounts and usefully understanding science and technology in society.

While it raises concerns for how both artifacts and knowledge are constructed by social actors, STS does not leave considerations of the material world behind. STS utilizes a materialist ontology in that it explores how scientists and engineers "attempt to construct stable structures and networks" through the use of the material world; knowledge and artifacts are indeed human products, marked by the circumstances of their production, both social and material (Sismondo, 2004, p. 10-11).

STS as a field has fought against the currents of both technological determinism and social constructionism that at times preoccupy many in the social sciences. Theories of sociotechnical systems, including actor-network theory, are both informed by these currents, and as STS does in general, push against them. These perspectives are all relevant to explore when applying STS to new/digital media.

Technological Determinism

Livingstone (2002) contends, "Despite a range of critiques, technological determinism remains alive and well and, whether in academic, public or policy forums, significant social changes are being attributed to technological innovation" (p. 18). This determinism is in large part a legacy of the positivist tradition that has supported the scientific method since the Enlightenment; causal relationships are prized for their authority, and technology, itself the result of science, inherits this authority and enables subsequent social action and behavior (Sismondo, 2004, 9). Technological determinism, then, posits that forms of emergent technological artifacts can predict, and in many fundamental ways shape, social behaviors and structures (Jackson, Poole, and Kuhn, 2002). Some new media research attempts to avoid a position of determinism by ignoring or marginalizing the details of the technical on principle (thus leaving those details in a black box), and it is a challenge for researchers interested in a more balanced perspective to understand both the social landscape and the technological specificities of new media (Livingstone, 2002).

More fundamentally, Leonardi and Barely (2008) claim that misunderstandings of technological determinism have left the materiality of information technology "grossly under-theorized" (p. 161). They write that "the legacy of materialistic determination [has] acquire[d] a kind of moral authority," and that the continuation of this legacy is the result of a common conflation between determinism and materiality on the one hand, and agency and social ideals on the other:

> Although the distinction between determinism and voluntarism [agency] is orthogonal to the distinction between materialism and idealism, social scientists frequently write as if materialism implies determinism and idealism implies voluntarism. This is simply not the case. (Leonardi & Barely, 2008, p. 160-161)

These scholars go on to outline how this determinism developed from the perspectives of Marxist theory, contingency theory, organizational theory, and media richness theory, while offering examples of how "attending to agency and social dynamics is not incompatible with an appreciation for material constraints and affordances" (Leonardi &

Barley, 2008, p. 163). In the end, a turn to a sociotechnical approach grounded in STS could allow researchers to both broadly and precisely understand many of our daily interactions with information technology and new media.

Social Construction of Technology

A counter position to material determinism that emerged during the mid-twentieth century is social constructionism, which "emphasizes the role of humans in actively using symbolic resources to objectify, circulate, and interpret the meaningfulness of their environments and their existence" (Lindlof & Taylor, 2011, p. 45). In the context of STS, the constructionism perspective maintains that technological elements alone cannot explain the social elements of an environment, but rather that technical and social factors are "intimately interconnected" (Jackson, Poole, & Kuhn, 2002, p. 237).

The Social Construction of Technology (SCOT), first proposed by Pinch and Bijker (1987), crystalizes many of the major tenets of constructionism and has become a key perspective for STS. SCOT challenges the notion that technologies have essential features that then can have systematic effects on the social world; instead, technologies are viewed as constantly in flux because of their continuous engagement with social communities (Kline & Pinch, 1999; Jackson, Pool, & Kuhn, 2002). Technologies have *interpretive flexibility*; their trajectories should be seen "as the result of rhetorical operations, defining the users of artifacts, their uses, and the problems that particular designs solve" (Sismondo, 2004, p. 81). As demonstrated in Pinch and Bijker's (1987) seminal analysis of the modern bicycle design, but also applied by Wu (2010) to many of the major technological developments in mass communication since the telegraph, the success of a technology often depends on the strength of its chief proponents, whether that strength is economic, political, or social. Stabilization or *closure* of a technology can then occur based on these social factors, but the interpretive flexibility of the artifact is always open to resistance and further influence from any social group (engineers, advertisers, consumers, etc.) that constructs meaning around the technology (Kline & Pinch, 1999).

Sismondo (2004) writes, "To accept that technologies do not have essences is to pull the rug out from under technological determinism" (p. 83). Kline and Pinch (1999) add that the anti-essentialist position of SCOT emphasizes "the dangers of the analyst assuming a taken-for-granted bedrock of a technical realm that sets the meaning of an artifact for all spaces, times, and communities" (p. 114). Ultimately, studies employing SCOT assumptions have been criticized for focusing too much on the design stage of technologies and not enough on their adopted uses and subsequent meanings, a critique that sociotechnical systems theory and actor-network theory attempt to remedy.

Sociotechnical Systems

Bijker (1995b) describes social science's interest in technology in the twentieth century as a swinging pendulum, moving back and forth (often too far) between technological determinism and social constructionism. The more recent turn to understanding "sociotechnical ensembles," Bijker argues, has greatly reduced these swings while offering a nuanced perspective on technology and society. Nearing the end of the century, he writes, "Technology [has] recaptured some of its obduracy without completely losing its socially shaped character" (Bijker, 1995b, p. 254).

Though its genesis precedes the formalization of STS, the personal computing revolution, and many of the digital tools that are now integrated into our daily lives, the concept of sociotechnical systems is strongly linked to STS and informed by new/digital media theory. At the heart of the concept is a desire to reconcile the divide between the social and the technical that resulted more broadly from the Industrial Revolution, and more specifically from Shannon's information theory. A sociotechnical systems perspective seeks to understand how heterogeneous ensembles of technical, social, political, and economic elements work together to solve problems.

The original concept of sociotechnical systems can be traced to the Tavistock Institute in London during the mid-twentieth century, where Trist and Bamforth (1951) observed the behavior and organization of the British coal mining industry, concluding that design principles need to consider both the worker and the work task in order to maximize the satisfaction of the former and the efficiency of the latter. The theories for organizations they developed highlighted concerns for the health of a system as a whole (Pasmore, 1988). Such theories, though, conflicted with the prevailing spirit of Taylorism that permeated industry in the early-twentieth century and reinforced ideas of technological determinism. Taylorism operated on the premise of closed-systems with managerial control over the work design and minimal concern for the humanity of the workers (Braverman, 1974). In addition, information theory's premise of controlling closed systems of message transmission to maximize efficiency served to separate concerns for the social and the technical aspects of communication further.

The development of sociotechnical systems theory necessitated a shift to open-systems thinking, as to understand the complexity of real work situations, one needs to consider relations within the system as well as relations to the environment. Emery and Trist (1965) looked to the behavioral sciences for inspiration, where they found researchers differentiating “living” organisms and organizations that import necessary materials from the environment for survival and stability, from inanimate objects that lack such agency.[11] Trist (1959) and the early sociotechnical systems theorists used this model of interaction with the environment but modified the assumption that only living organisms are agents in the system; by elevating technology to a more important, though not deterministic, role, they advanced the notion that “open system and socio-technical thinking imply each other,” an expansion of ideas on social systems put forth by Weber, Parson, and Merton (p. 43-44). Thus, an early definition from industrial and organizational studies states:

> The sociotechnical systems perspective considers every organization to be made up of people (the social system) using tools, techniques and knowledge (the technical system) to produce goods or services valued by customers (who are part of the organization’s external environment). (Pasmore, 1988, p. 1)

The concept of sociotechnical systems informs theory and practice in organizational communication, knowledge sharing, and business (Choi, Kang, & Lee, 2008; Gregoriades & Sutcliffe, 2008; Preda, 2006), as well as engineering (Johnson & Wetmore, 2008) and information technology (Benoit-Barne, 2007; Bryl, Giorgini, & Mylopoulos, 2009). Johnson and Wetmore (2008) succinctly sum up the concept:

> Sociotechnical systems acknowledge that attempts to understand a device or a social practice (institution, relationship, etc.) as an independent entity are misleading. To treat either as a separate unit is to abstract it from reality. … A focus on sociotechnical systems helps us see the ways in which artifacts, social practices, social relationships, systems of knowledge, institutions, and so on are bound together and interact with each other in complex ways. (p. 574)

They go on to describe the ways in which “engineering is a moral and political endeavor” where social practices, the distribution of benefits and burdens, and mechanisms for

[11] It is interesting to note that the literally “biological” antecedents to sociotechnical systems theory—research into ant colonies and locust swarms—are similar to those of collective intelligence, as described later in Chapter VI.

freedom and control are as much the product of the technologist's work as the technological artifacts produced (Johnson & Wetmore, 2008, p. 575).

Emery and Trist's (1965) early emphasis on open systems that consider both worker and work, developed more theoretically in actor-network theory and Latour's (1991) chains of association and social assemblages, have had important implications for thinking around issues of control and technology in the information society/economy. Boltanski and Chiapello (2005) have argued that by recognizing and acknowledging the related social and technical spheres of the information society, capitalists have been empowered to yield a new type of control in the economy, one centered on flexibility and communication. Fisher (2010) has traced this same argument through the modern discourse of a sociotechnical culture, showing how media such as *Wired* magazine prominently offer the promises of techno-culture without full exploration of the consequences of such a culture. Zittrain (2006) remains pessimistically hopeful that the open systems structures initiated by the convergence of the information revolution and the computer revolution will be more widely adopted and fostered in modern society, while Wu (2010) presents a bleak outlook for open systems, documenting the cycle that all modern communication technologies have followed from open development to monopolistic control.

Those devoted to studying the relationships between users and information communication technologies (ICTs) have established the field of social informatics, which Kling (1999), an early and vocal proponent of the field, defines as "the interdisciplinary study of the design(s), uses, and consequences of information technology that takes into account their interaction with institutional and cultural contexts” (p. 1). Bradley (2006) simplifies this definition to “the prerequisites, the applications, and the impact of ICTs on humans” (p. 72). Social informatics is a transdisciplinary field composed of scholars from computer science, information systems, new media studies, sociology, and communications (Sawyer & Tapia, 2007), as well as a conceptual area of research “defined by its topic (and fundamental questions about it) rather than by a family of methods, much like the fields of urban studies or gerontology” (Kling, 2000, p. 218). The foundation for social informatics is the recognition that ICTs are “inherently sociotechnical” (Sawyer & Tapia, 2007, p. 264) and that technology and social use are not examined separately, but seen to “co-constitute” each other (Kling, 2000, p. 220). Research centers on social informatics have been established at Indiana University, Bloomington, the University of California, Irvine, Syracuse University, and the London School of Economics, to name but a handful of locations.

More philosophically, Ropohl (1999) speculates on the importance of the sociotechnical perspective, as well as reasons why the concept of sociotechnical systems remains unsettling for some. Technological innovation can both replace human functions and add novel new functions that are unfeasible to humans:

> The states of the human subsystems and the characteristics of socio-technical relations are changing. Every invention is an intervention, an intervention into nature and society. That is the reason why technical development is equivalent to social change. (Ropohl, 1999, p. 69)

The transfer of human functions to more efficient technologies—the “loss” of these functions one might say—can be anxiety provoking in numerous respects (i.e. economically, socially, emotionally), and as technology becomes socialized, functions we lose to technology become alien. Ropohl (1999) uses the example of the pocket calculator: once a common ability amongst the educated, the calculator has given to everyone the ability to calculate the square root of a number, even if the math behind the calculation was never learned. The externality of this once human function can serve to alienate the user, especially as the ability to set goals, one of the most basic and still (almost) exclusively human functions, becomes farther abstracted into the interfaces of technology:

> The inconvenience of alienation results from the sociotechnical division of labor; it cannot be abolished, but it may be relieved by technological enlightenment, which aims to accompany the appropriation of artifacts by an appropriation of the appropriate understandings. (Ropohl, 1999, p. 70)

Indeed, the development of new understandings of technology and society necessitate the openness to new and often unfamiliar perspectives.

The sociotechnical systems perspective is particularly useful for studying digital media because it integrates the attributes of variability, cultural transcoding, and programmability that, as previously established, are central to their nature. By emphasizing relationships between agents, as well as the complex, multifaceted character of agents in an open environment, the approach avoids essentialist thinking that often produces dichotomous analysis—social vs. technical, material vs. semiotic, human vs. non-human, agency vs. determinism. Previous research into online platforms, including Wikipedia, have tended to use traditional approaches that highlight users or technologies but fail to consider the “dual nature” of new media that Manovich (2001), Galloway (2004), and Lessig (2006) emphasize. The sociotechnical systems perspective offers a framework from which to question and understand the simultaneous openness and control in a system like Wikipedia.

Actor-Network Theory

Sometimes the topic of controversy for its unconventional twist on social theory, actor-network theory (ANT) has become “the most successful of STS’s theoretical achievements so far” (Sismondo, 2004, p. 74) and “the field’s leading research orientation” (Fuller, 2006, p. 58). In line with the general perspective of STS, ANT holds that “science and technology are done in rich contexts that include material circumstances, social ties, established practices, and bodies of knowledge” (Sismondo, 2004, p. 69). Largely established by the work of Callon (1986), Latour (1987), and Law (1987), ANT is grounded in an ontology of relational materialism, where actors are defined by their place in a network with other actors. These networks are constantly in flux, as actors are continuously rearranging the network’s configuration by influencing other actors. Through this act of influencing, the meaning and power of an actor is *translated*, as it derives its meaning and power from its position in the network. ANT is also concerned with the translation of interpretation to fact, especially in the context of science, and understands the durability of facts as a contingency of the network; in other words, the scientific facts that have come to dominate our understanding of the world do not emerge a priori from nature, but rather are interpretations of phenomena made stable over time by actors in a network of meaning. Thus, facts themselves become material artifacts in the network, influencing actors and open to further translation by other actors. Overall, ANT strives to understand this making and remaking of a network and how it can create stable realities.

ANT argues that networks are composed of both human and nonhuman actors, and that no fundamental distinctions need to be made between the two. This controversial principle for relational analysis breaks with the tradition of Western sociology and offers a generalized symmetry that provides an "explanation of the development of sociotechnical ensembles involving neither technical nor social reductionism" (Bijker, 1995, p. 251). Sismondo (2004) adds:

> Representing both human and nonhuman actors, and treating them in the same relational terms, is one way of prompting full analyses, analyses that do not discriminate against any part of the ecologies of scientific facts and technological objects. It does not privilege any particular set of variables, because every variable (or set of actors) depends upon others. (p. 69)

Latour (1991) claims that "in order to understand domination we have to turn away from an exclusive concern with social relations and weave them into a fabric that includes non-human actants, actants that offer the possibility of holding society together as a durable whole" (p. 103). Although we are sometimes faced with experiences that seem purely social (human) or purely technical (nonhuman) in nature at the immediate level, these episodes are in fact two parts of a longer sequence of associations and meanings in our network, for which Latour (1991) comments: "power is not a property of any one of those elements but of a chain" (p. 110).

Two examples highlight the codependency Latour (1991, 1999) describes: the hotel room key and the speed bump. The pre-digital hotel room key commonly found in Europe would be bound to a weight or object of some sort in order to remind guests to return the key before leaving the premises. Where verbal and written methods of communication (reminders and signage), as well as social/moral obligation, usually failed in securing compliance, the weight would succeed at this seemingly small task; the technology here is inscribed with the social meanings and relations of the situation. In fact, technology can bridge and negotiate disparate social meanings to create order, as in the case of a speed bump near a school. The social meanings inscribed in the speed bump include “slow down so as not to endanger students,” as well as “slow down and protect your car’s suspension” (Latour, 1999). These goals likely register with varying importance in different drivers’ minds, but in the technology of the speed bump—a translation of these social expressions—is the power to change behavior in a manner often more effective than that of purely human relations (moral duty, concern for others, self-interest, etc.). From these examples, Latour (1991) concludes, “We might call technology the moment when social assemblages gain stability by aligning actors and observers. Society and technology are not two ontologically distinct entities but more like phases of the same essential action” (p. 129).

Much of ANT’s success as a social theory, as well as a theory for STS, has been an insistence on relational materiality, intuitively explaining the actions, forces, and interests that shape scientific fact and technological development. Numerous criticisms of the theory have been presented, however, including its failure to incorporate cultural networks and social values like trust into its analysis; its elevation of nonhuman actors (or conversely, its demotion of human actors) in the name of symmetry; and its tendency to privilege the agency of humans over nonhumans, despite its tenet of relational uniformity between all actors (Sismondo, 2004). For research in the field of communication, though, Lindlof and Taylor (2011) cite ANT as a useful perspective, as communication is undeniably bound up in the translations of actors in a network. For this project, perspectives and viewpoints informed by ANT and STS allow the network of social and technical actors that make up Wikipedia to be more thoroughly interrogated and their co-constitution to be more richly explored.

The Political Economy of Online Communication

Political economy has been a foundational perspective in the social sciences since the eighteenth and nineteenth centuries, when Smith (1759, 1776) and Ricardo (1817) most notably wrote about the nature of economics and the role of markets in society. Critical political economy, though, starting with Marx (1867), has sought to more thoroughly and revealingly question the theoretical and material contradictions of economic organization that have been a driving force for history. Today, critical political economy informs work on markets and labor, corporate and government structures, and social disparities of class, race, and gender across a number of disciplines, including communication studies. From the early work of the Frankfurt School on the culture industries (for example, Horkheimer & Adorno, 1947), through Murdock and Golding's (1973) work on industry structures and Smythe's (1977) work on the audience commodity, to the more recent work of Wasko (2004) on modern media systems and Fuchs (2011a) on ICTs, critical political economy has been an important framework for understanding relationships between the social, economic, and political in mass communication. With the current dominance of digital media in everyday communication, including the diffusion of labor through online networks, it is imperative to continue examining these relationships and their meaning for a site like Wikipedia.

The Political Economy of Communication

Quite contrary to Shannon's Mathematical Theory of Communication, Mosco (2009) defines communication as "a social process of exchange, whose product is the mark or embodiment of a social relationship" (p. 67). The author then defines the political economy of communication as "the study of the social relations, particularly the power relations, that mutually constitute the production, distribution, and consumption of resources" (Mosco, 2009, p. 2). In her review of the approach, Wasko (2004) argues that these relations are explored by examining various tendencies of the media industries, including commodification, diversification, horizontal and vertical integration, synergy, and market concentration. By understanding the power relations that drive the media industries and the mechanisms of domination commanded by its controlling class, we can understand the "pervasive ideological character" of mass media commodities (Schiller, 1989, p. 33). Informed by the political economy of communication, Herman and Chomsky (1988) explore such hegemony in mainstream journalism, while Jhally (1990, 2006) and Klein (2000) look at advertising and consumer culture. Though not dismissing such message-based analysis, others in the field continue to keep the political and economic structures of the mass media central to their work (Bagdikian, 2004; McChesney, 2008; Meehan & Torre, 2011; Wasko, 2003).

Many critical scholars have highlighted the nuanced analysis that can be achieved by using the political economy of communication in conjunction with other perspectives, including cultural studies (Kellner, 2009; Murdock, 1995) and feminist studies (Meehan & Riordan, 2002; Steeves & Wasko, 2002). STS offers a more recent ally for political economy, as both perspectives start from a realist epistemology, both concentrate on the relationship between knowledge and social practice, and both are interested in issues of democracy (Mosco, 2009). As Fuchs (2011a) argues, both perspectives utilize dialectical reasoning to avoid technological and social determinism, but Mosco (2009) points out:

> STS moves beyond even the most ambitious definition of political economy, which calls for the study of control and survival in social or even organic life. STS does not stop at social life because of the centrality of organic life, but it also wishes to energize technology. The latter is not just an inert mass, the computer

> on the desk, but a force that grows, retreats, and otherwise interacts with non-technological actors in its network. (p. 235)

Elsewhere, Mosco (2004) writes, “We continue to compartmentalize politics and rhetoric as forces external to an entirely different process, which we call science. But [STS] scholars compel us to examine how politics and rhetoric are constitutive of the scientific enterprise” (p. 13). As our environments become more inundated with ICTs, the need for both political economy and STS in a critical assessment of technology becomes obvious.

Online Communication

Well before the mass adoption of online communication in developed areas of the globe, Bell (1973) and others predicted that technology would play a major role in the shift from an industrial to a post-industrial society, one marked by a service-oriented economy and a controlling class of technical elites. The Internet and online communication have certainly played the role of society-changing technology, but contrary to Bell, some digital scholars perceive its influence as ultimately empowering and democratizing (Benkler, 2006; Rheingold, 1993). With knowledge of traditional media and an analytic approach, political economists looking at this development ask “to what extent can the emerging communication technological revolution, particularly the Internet, override the antidemocratic implications of the media marketplace and foster more democratic media and a more democratic political culture?” (McChesney, 2008, p. 356).

Although born of government funding and academic initiative, the Internet became fully commercialized in 1995 (Abbate, 1999). Thus, much political economic inquiry into the Internet has looked at issues of ownership, control, and regulation for online communication. Vaidhyanathan (2011) and Lee (2010) examine Google’s dominance of online search, as well as the company's conglomerating tendencies and expanding ownership of content. Wasko and Erickson (2009) dig into one of Google's holdings, YouTube, to uncover how the company is monetizing free, user-generated content. Soderberg (2002) and Tian (2009) explore law and policy around both copyrighted and open-source/open-content materials online, while Mendelsohn (2011) and Starosielski (2011) track who controls the material infrastructure (cables and routing stations) that enables worldwide communication on the Internet.

Beyond traditional paradigms of media ownership and control, though, political economists are also investigating the mechanisms for information surveillance in interactive digital media and the possible implications of this surveillance for society. Citing social networking sites like Facebook, MySpace, Flickr, and Twitter, Fuchs (2011b, 2012) argues that users and their data are constantly being commoditized by Web 2.0 applications, sometimes unknowingly or unwillingly, in the name of providing a personalized online experience. Though surveillance and privacy are certainly not new concerns for society, the author contends that the hyper-aggregation of data by companies like Facebook and Google favors market concentration which poses new barriers to participatory democracy and an open Internet. Fuchs et al. (2011) survey additional issues pertaining to surveillance and privacy on the Internet, including self-disclosure of information, the right to access personal data, file-sharing, protest and social movements, and international cyberlaws.

Digital Labor

Mosco (2011) writes, "If, as Dallas Smythe (1977) famously remarked, communication is the blind spot of western Marxism, then labor remains a blind spot of western communication studies, including the political economy tradition" (p. 358). From Marx (1867) through Braverman (1974), critical political economists have explored the exploitation of physical labor at the hands of capitalism, but more work is needed within communication studies to understand the relationship between online intellectual labor and the commercial interests of the Internet. Understanding intellectual labor is particularly key to an analysis of Wikipedia, as nearly all of the work done on the site is furnished by volunteers.

Hardt and Negri (2004) popularized the notion of *immaterial labor*, which they define as "labor that creates immaterial products, such as knowledge, information, communication, a relationship, or an emotional response (p. 108). Immaterial labor in the form of communication is necessary for the function of society, and is therefore a public "common" that would ideally be cooperatively owned by society; however, under capitalism this labor is exploited and converted to surplus value by the corporate class (Hardt and Negri, 2004). More so than the traditional audiences of legacy media, users of digital media are often content producers as well as consumers, an added dynamic of media convergence (Bruns, 2008; Jenkins, 2006). Lessig (2006) and Zittrain (2008), among others, argue that the Internet is organically a communications common and the virtues of free, open information and unrestricted participation drive its development. As previously discussed, though, many Internet sites have become what Andrejevic (2007) calls a "digital enclosure," where communication data and personal data become the private property of the site, only to be fed back to the users in the form of advertising.

Popular culture scholars like Jenkins (2006) have done well in documenting much of the digital labor pervading the Internet, even suggesting that fan communities and content producers use their collective intelligence to push back against the traditional culture industries and ultimately have a say in the market economy. Terranova (2004) also identifies collective intelligence as a key trait of online immaterial labor, though her analysis is from a more critical political economic perspective:

> As a collective quality of the labor force, immaterial labor can be understood to pervade the social body with different degrees of intensity. This intensity is produced by the processes of 'channeling' of the capitalist formation which distributes value according to its logic of profit … music, fashion, and information are all produced collectively but are selectively compensated. Only some companies are picked up by corporate distribution chains in the case of fashion and music; only a few sites are invested in by venture capital. However it is a form of collective cultural labor which makes these products possible even though the profit is disproportionately appropriated by established corporations. (p. 83-83)

Both Terranova (2004) and Murdock (2010) argue that the digital economy is really a moral economy where state, capital, and civil interests converge; the result is a space that exhibits both a market-driven compulsion and a gift economy reminiscent to tribal cultures. To understand digital labor, then, the political economist must investigate the relationship between these contrasting logics and the power dynamics that emerge. Ultimately, Murdock (2010) finds capital too exploitative for this relationship and, similarly to Lessig (2004), calls for the establishment of a digital commons to preserve the openness of digital labor and to keep public and cultural knowledge and artifacts free and accessible.[12]

Though a plethora of research on Wikipedia has been published over the past decade, with the following section presenting major trends and findings from that work, few studies consider the project from a political economic perspective. For this study, the

---

[12] Wikimedia Commons is one such effort. Overseen by the WMF, Wikimedia Commons contains over 12 million media files used in the various Wikimedia projects. These are fair use media objects, often carrying a Creative Commons or GNU Free Documentation license.

preceding issues and ideas are crucial to fully understanding the network of communication, collaboration, and labor that propel the site.

Wikipedia and Online Participation

Though Wikipedia rarely made the pages of leading academic journals during its formative years, the past five years have seen an explosion in research on the site and online collaboration more generally. This work builds from the solid base of research on the Internet, online networks, and user participation established by computer scientists and social scientists over the past two decades. Relevant studies on Wikipedia and online participation will be discussed and cited throughout this project, but some foundational literature is offered here to help frame the need and significance of the present work.

Motivations to Contribute Online

Motivation has offered a rich avenue of inquiry for social scientists trying to understand the digital phenomenon of voluntary engagement and contribution to online platforms, including Wikipedia. Indeed, many argue that the development of the Internet before the emergence of the World Wide Web (and through the early days of the Web, and even today in certain ways) was a "labor of love," a feat of expert and amateur passion and devotion (Benkler, 2006; Zittrain, 2008) that was subsequently exploited by technology corporations when the personal computer revolution swept the Western world (Coupland, 1995; Wu, 2010).

The voluntary dimension of digital production has been the subject of much research over the past decade in numerous settings (Brabham, 2010; Utz, 2009; Wasko & Faraj, 2005). Survey methods have been employed to probe the motivations of open-source software (OSS) developers, demonstrating that motivational processes are similar to other social communities and “can be explained within existing social psychological theories” (Hertel, Niedner, & Herrmann, 2003, p. 1174) like intrinsic and extrinsic motivation (Osterloh & Rota, 2004). Intrinsic motivations for participating in OSS projects include enjoyment, the public display of one’s ability, personal obligation to peers, and the belief in the importance of reciprocity, while extrinsic motivations are generally focused on concerns for one’s reputation and the personal needs that can be satisfied by the project. Osterloh and Rota (2004) argue that while both types of motivation should be expected in collaborative environments, without intrinsically motivated contributors “cooperation in open-source software is not sustainable” (Osterloh & Rota, 2004). Further, the social psychological concepts of social loafing and social caution can be applied to online collaborative projects, with abstract group affiliation potentially having a negative effect on information sharing behavior, though the context of the affiliation is key (Cheshire & Antin, 2010). Oreg and Nov (2008) explore the variable of context deeper to find that contributors to OSS projects, where expertise is more highly prized and work undergoes a stricter peer-review, are more highly motivated by self-development and reputation-building, while contributors to open-source content projects (like Wikipedia), with greater acceptance of amateurs and lower barriers to entry, work with more altruistic motivations.

A number of recent studies have probed deeper into the motivations of Wikipedia editors by surveying either the general population or specific groups of contributors. In an early study, Nov (2007) found that “fun” was users’ highest ranked motivation and was strongly correlated with the number of hours per week spent contributing to Wikipedia, while “ideological” motivations, though ranked second, did not correlate to contribution level. More specifically considering Wikipedia administrators, Baytiyeh and Pfaffman (2010) discovered both intrinsic and extrinsic motivations for these highest-privilege users; intrinsic motivations like the desire to learn and create were strongest, along with extrinsic motivations to create and connect with a community, though the desire to make Wikipedia a respected online source of information was strong for a smaller subset of administrators. On the opposite end of the user spectrum, Shachaf and Hara (2010) found that Wikipedia trolls (or vandals) share many of the motivations of hackers, including intrinsic (excitement, intellectual curiosity, challenge) and extrinsic (prestige and power) reasons for their harmful actions.

Overall, the literature indicates that Wikipedia editors are motivated less by external concerns for reputation and status than many other online platforms. Instead, contributors devote their time, energy, and expertise to Wikipedia out of enjoyment, satisfaction, and altruism. The Wikimedia Foundation is itself invested in learning more about who contributes and why they contribute. At the beginning of this study, the organization was collecting responses for its "Wikipedia Editors Survey 2011" both to update its demographic data on editors and to understand patterns of communication and potential discrimination in the system (Wikimedia Meta-Wiki, 2011d). And in more focused efforts, Wikipedia's parent foundation is exploring barriers to participation, including why scientists, academics, and experts do (or do not) participate in content creation on Wikipedia (Wikimedia Meta-Wiki, 2011d). These actions clearly signal the organization's awareness that the direction of the project's evolution is tied to who contributes and what they contribute.

The Digital Divide and Online Content

Many of the early political and theoretical discussions around information bias on the Internet have focused on issues of digital divide, where populations of people are barred from online participation due to physical, economic, or social barriers (Castells, 2001). This digital divide is generally conceptualized macroscopically as inequalities of digital access between the developed and undeveloped (and under-developed) nations and regions of the globe (Norris, 2001). But other discussions focus on those of us who are connected and depend on networked information in our daily lives. In certain contexts, we should not be surprised to find bias favoring specific information. Response bias in eBay seller ratings (Dellarocas & Wood, 2008), corporately-crafted "anonymous" blog comments (Barbaro, 2006), and questionably-sponsored health information (Stvilia, Mon, & Yi, 2009) have all come to light in the Net era, but in a sense, these biases fall under the "caveat emptor" spirit of modern capitalist society.

Others have focused more closely on how Internet architecture and protocol can create information bias, another form of digital divide. Online search engines are a less obvious harbor for information bias than user-generated content like online consumer ratings. Google, the highest traffic site on the Internet, uses an advanced PageRank algorithm to deliver search results including sponsored links, which appear at the top of the page but are clearly separated from the non-sponsored links and delineated as "Ads." As Segev (2010) argues, though, sites like Google not only facilitate information distribution but act as the gatekeepers of online information. Segev (2010) looked closely at search engine technology, and specifically Google, to find that algorithms and customization can be exploited by advanced users, large corporations, and even nation-states to organize information in a manner that intensifies inequalities and supports commercial Western agendas. Gillespie (2011) has similarly examined the technology that drives Twitter's Trends list, concluding we must "recognize that these algorithms are not neutral, and that they encode political choices, and that they frame information in a particular way."

When looking at Wikipedia content, many of the same concerns regarding inequality and bias are found. While some researchers have considered the entirety of Wikipedia's content in their studies (Kittur & Chi, 2009; Wilkinson & Huberman, 2007), an approach made feasible by freely available *data dumps*,[13] others have taken a more focused approach through sampling (Goldspink, 2010; Halavais & Lackaff, 2008) or purposive case studies (Kane, Majchrzak, Johnson, & Chenisern, 2009; Konieczny, 2009). Each method is useful to elucidate what is and is not present in the online encyclopedia. And as Wikipedia is an active repository of information, changing from moment to moment with the real-time actions of its users, those studying its content need to either negotiate the limitations of a snapshot approach, or build the variable of time into their methodology (Kimmerle, Moskaliuk, Harrer, & Cress, 2010; Tkacz, 2010).

Building from the work of Tankard and Royal (2005) on the completeness of information on the World Wide Web, Royal and Kapila (2009) offer perhaps the most pointed research to date on Wikipedia content and issues of bias and representation. By systematically searching for gaps in Wikipedia content through both targeted searches and random sample searches, the authors conclude that the site succeeds in its "broad coverage" of encyclopedic information. However, Royal and Kapila (2009) found a

[13] Database dumps are complete copies of the tables in a database, which allow analysis without altering the live version of the data. As Wikipedia content is published with Creative Commons and GNU Free Documentation licenses, data dumps are freely available for research and personal use.

number of significant biases: more common or popular terms have the most detailed coverage; coverage of large nations is positively correlated with national population; and companies with larger revenues and assets are more likely to be covered in-depth. Halavais and Lackaff's (2008) study of Wikipedia content also showed an emphasis on popular culture and recent events, while subjects like "physics" are underrepresented compared to traditional printed encyclopedias. Kittur, Chi, and Suh (2009) further confirmed this emphasis on culture, though their research indicates that the natural and physical sciences, philosophy and thinking, and mathematics and logic are among the fastest growing areas of Wikipedia's content. Collectively, these inquiries suggest that "Wikipedia is more a socially produced document than a value-free information source," and its content is a volatile, dynamic, and ongoing reflection of its editors' interests (Royal & Kapila, 2009, p. 146).

Governance and Control on Wikipedia

In addition to research on motivation, access, and content, a number of studies have looked into issues of governance and control on Wikipedia. Although the technical infrastructure of the site is built to support and perhaps encourage an equal distribution of power on the site, previous research into the site's management indicates Wikipedia is not a land of "anything goes." The popular press has covered efforts by the site to reduce vandalism through a layer of editorial review (Manjoo, 2009; Sutter, 2009), a tightening of control cited as a possible reason for the recent dip in the number of active editors (Angwin & Fowler, 2009; Johnson, 2009; Ortega, 2009). A number of regulations are already in place to prevent the open editing of certain articles and pages, such as the site's disclaimers and pages that have suffered large amounts of vandalism. Editing wars can also cause temporary restrictions to editing, and Ayers, Matthews, and Yates (2008) point out that these wars can happen anywhere, from the article on George W. Bush to that of film-star Burt Reynolds.

Academic studies into Wikipedia's governance generally highlight how order is maintained not through particular actors, but through established procedures and norms. Konieczny (2009) tested whether Wikipedia's evolution can be defined by Michels' Iron Law of Oligopoly, which predicts that the everyday operations of any organization cannot be run by a mass of members, and ultimately control falls into the hands of the few. Through exploring a particular WikiProject on information validation, he concluded:

> There are few indicators of an oligarchy having power on Wikipedia, and few trends of a change in this situation. The high level of empowerment of individual Wikipedia editors with regard to policy making, the ease of communication, and the high dedication to ideals of contributors succeed in making Wikipedia an atypical organization, quite resilient to the Iron Law. (p. 189)

Butler, Joyce, and Pike (2008) support this assertion, though they emphasize that instead of oligarchy, control becomes encapsulated in a wide variety of structures, policies, and procedures that guide involvement; a virtual "bureaucracy" emerges that maintains order on the site.

Other work considers control on Wikipedia through the framework of commons governance. Viégas, Wattenberg and McKeon (2007) explain that "peer production depends on individual action that is self-selected and decentralized rather than hierarchically assigned. Individuals make their own choices with regard to resources managed as a commons." The need for quality standards and quality control largely dictate this commons governance, though interviewing Wikipedians with various levels of responsibility revealed that policies and procedures are only as good as those who maintain them. Forte, Larco, and Bruckman (2009) argue, “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” (p. 71).

Conclusion

The study of new/digital media, as is the study of most communication, is an interdisciplinary enterprise. This chapter reviewed major theories and concepts from the research traditions that inform the present study and serve as a foundation from which to construct appropriate research questions, choose appropriate methodologies, and ultimately interpret findings. The following chapter discusses in more detail the first two of these: research questions and methodologies.

CHAPTER III

RESEARCH QUESTIONS AND METHODOLOGY

Research Questions

As traced in the previous chapter, a discussion of digital media calls forth a confluence of theories and ideas from many different disciplines, including computer science and mathematics, computer engineering, sociology and labor studies, philosophy, and mass communication. Of course, our everyday experience of digital media, whether it be purchasing or listening to a song on iTunes, sending a text on a smartphone, or updating a Facebook status, does not necessitate such a deep and heterogeneous inquiry into the "hows" and "whys" of these tools and platforms. More and more, though, everyday experiences—as well as extraordinary experiences like the 2011 civil protests in Northern Africa—show us how deeply enmeshed "the social" and "the technical" are becoming in our communication practices. It is vital, then, to understand both the technical infrastructures of our media systems and the social forces of command and control, be they bureaucratic like the Stop Online Piracy Act (SOPA) in the U.S.[14] or autocratic like the shutdown of the Internet in Egypt,[15] that dictate their use.

A flood of research on Wikipedia has appeared in recent years, with more to come as the site matures and looks for new areas for growth and expansion. Most of this scholarship, however, considers either the social structures that have developed in the

[14] The Stop Online Privacy Act is a U.S. House of Representatives bill introduced in 2011 that proposes to expand enforcement of online copyright law. A similar bill, the Protect IP Act (PIPA) was introduced in the U.S. Senate the same year. At the time of this writing, floor votes on each bill had been postponed, likely in response to the January 18, 2012 online protests by major websites like Google and Wikipedia (Kane, 2012).

[15] Reports indicate that on January 27, 2011, the Egyptian government attempted to block internet access within its borders in response to organized street protests , successfully shutting down up to 88 percent of connections (Kanalley, 2011; Williams, 2011).

project, or trends in content creation or user activity. This dissertation research considers issues of social and technical convergence on Wikipedia that are largely unaddressed in the research literature, and the project is guided by the following research questions:

RQ1: What is the historical context for the technical development of Wikipedia? What actors and forces have shaped its code, community, and infrastructure?

RQ2: In what ways is Wikipedia a sociotechnical system? What roles do the social and technical infrastructures of the site play in its collaboration and conflict? How does science and technology studies (STS) contribute to an understanding of the site's performance?

RQ3: In what ways is Wikipedia a system of collective intelligence? How is this frame applied and understood by the mass media? The research community? The Wikipedia community? How do other concepts of intelligence inform this discussion?

RQ4: What can a social network perspective reveal about the collaboration between human and non-human actors on Wikipedia? How could this collaboration be visualized? What implications emerge from this picture?

Human Subjects Research and the Institutional Review Board

This study's methods for data collection involving contact with human subjects (interviewing) was reviewed by the Office for Protection of Human Subjects (OPHS) at the University of Oregon. The protocol (08262011.107) was determined to be exempt by OPHS on August 30, 2011, not requiring additional review unless the research continues beyond five years. In addition, a Research Clearance Form with all necessary signatures from the School of Journalism and Communication and the OPHS was filed at the Graduate School on October 13, 2011. Randall M. Livingstone was the primary and sole investigator for this study, and he completed the necessary Collaborative IRB Training Initiative (CITI) on August 10, 2011 (expiring two years later). All interactions with human subjects strictly adhered to the regulations and ethical considerations set forth by OPHS.

In line with the approved IRB protocol, interview subjects were emailed the study's Invitation Letter, listing the benefits and risks of the research, to obtain informed consent (Appendix XXX). The Invitation Letter was also posted on the researcher's Wikipedia user Talk page. Subjects were informed that by reading the letter and agreeing to be interviewed, they were providing informed consent, and that the interview would be archived and/or recorded (audio/video), when applicable. In addition, subjects were asked for permission to use their Wikipedia usernames in the dissertation and any subsequent publications; if a subject did not grant permission, a pseudonym (BotOp1, BotOp2, etc.) was created for use in the study.

Wikimedia Foundation Research Committee

In 2010 the Wikimedia Foundation Research Committee (RCom) was established to "help organize policies, practices, and priorities around Wikimedia-related research" (Wikimedia Meta-Wiki, 2012d). RCom's charter lists a multitude of more specific functions to aid and supervise researchers studying any of the WMF's projects, including Wikipedia. Most notably, these functions include "supporting the development of subject recruitment processes," "articulating and channeling requests for data and technical resources," and "setting guidelines for use of aggregation and research that singles out Wikimedians" (Wikimedia Meta-Wiki, 2012d). The committee currently consists of 14 members, including two employed by the Wikimedia Foundation. All members work in academia in some capacity and are experienced contributors to Wikimedia projects.

At the time this dissertation project was launched, RCom was still refining its policies, procedures, and documentation for research projects using data collected on its sites, and thus, the necessity for research approval by RCom was not clear. The WMF does maintain a "Notes on good practices on Wikipedia research" page that makes a distinction between "active" and "passive" research:

> Passive research … will be defined as a research method that does not require or require[s] a minimum interaction with Wikimedia or Wikimedians. Passive research is done on publicly available datasets and should require no authorization/approval from Wikimedians or the Foundation. Active research, on the other hand, requires the participation of Wikimedians or interactions with Wikimedians and will require some type of consent or authorization. (Wikimedia Meta-Wiki, 2011b)

Similar to the mission of Institutional Review Boards, RCom's intent is to minimize harm to human subjects, which in this venue often means reducing or eliminating "annoyance and disruption" to the user community (User:Ironholds, personal communication, May 31, 2011). The Wikimedia community has a number mechanisms for detecting active data collection on the site, including volunteer patrolling for solicitations on various Talk pages, reporting from individual users who are solicited, and monitoring of Recent Changes pages by either human users or bots for suspicious, repetitive edits to user Talk pages. When any of these activities is detected, the user associated with the actions is contacted regarding the situation and the account may be blocked until the situation is resolved (if the actions are coming from an unregistered IP address, the IP is likely to be blocked). Users are then directed to communicate with RCom to discuss their project and begin the approval process.

The process of gaining approval for active research on a Wikimedia site is much like the approval process for any action that may affect the community at large: formal documentation of the project must be made public, a waiting period for comments and discussion from RCom and the community must be established, and an overall positive consensus must be reached. Documentation must include a project summary, details on recruitment of participants, and the benefits of the project for Wikipedia. Prior approval from a researcher's IRB must also be posted, including protocol number and date of approval. In addition, the researcher is asked to state how the study's findings will be shared with the community and choose copyright options that are compatible with the Creative Commons CC-BY-SA 3.0 license under which the Wikimedia projects operate. An RCom member is assigned to each proposal and ultimately grants the go-ahead based on the consensus process. The RCom then serves as a resource for both the researcher and study participants to ask questions and receive support while the project progresses.

This dissertation project's research proposal, entitled "Understanding the Editor/Bot Relationship" was made public for comments on September 12, 2011 and was approved by Goran Milovanovic of the RCom (based on community consensus) on September 28, 2011 (Wikimedia Meta-Wiki, 2011c).

Methodology

Overview

There is a popular adage within the Wikipedia community: the site "works in practice, not in theory" (Ayers, Matthews, & Yates, 2008; Wikipedia, 2007b). Over the past decade, a broad array of methodologies have been used to shed light on the paradoxical nature of collaboration on Wikipedia, from discourse analysis to case studies, institutional analysis to ethnography, and surveys to content analysis. These methods have been grounded in diverse epistemologies and theories from communication studies research as well as many other academic fields. As Tkacz emphasizes, though, new approaches must continue to develop as “we need to continue exploring the *how* of collaboration” (cited in Walsh & Apperley, 2011).

This study employed two primary research methods—document analysis and interviewing—and one complimentary method—social network analysis—in order to address the “how” of collaboration between editors, developers, and automated tools of Wikipedia. Together, these methods allowed the researcher to triangulate data, contextualize information, and test existing theories. Discussions of these methods, as well as limitations of the study, follow.

Netnography/Ethnography

This study employs both qualitative and quantitative methods of inquiry to address the preceding research questions. Before describing these specific methodologies, however, it is important to note some of the broader research traditions from which this study drew inspiration for its approach, including digital methods and netnography.

In *The End of the Virtual*, Rogers (2009) calls for “a new era in Internet research”:

> The Internet is employed as a site of research for far more than *just* online culture. … The conceptual point of departure for the research program is the recognition that the Internet is not only an object of study, but also a source. … One is not so much researching the Internet, and its users, as studying culture and society *with the Internet* (p. 8, 29)

The approach here is to experiment with medium-specific methods and tools that are less concerned with early Internet online/offline distinctions and more attuned to understanding the often unstable relationships between actors and objects in the contemporary networked society. Groups like the Digital Methods Initiative (2012) have been established to explore and develop how methods like linking, searching, and networking can answer questions both quantitatively and qualitatively in the social sciences. In addition, algorithmic data collection and post-demographics[16] now enable new research questions on behaviors and relationships to be explored by the digital researcher, and extensive archival data on websites and in databases offers enhanced abilities to recognize patterns and predict future actions. In sum, digital methods should strive to alter, enhance, or even leave behind traditional social science methods, rather than merely importing them to the digital context (Rogers, 2009).

A recently documented digital method apropos to this study is netnography. Kozinets (2010) describes netnography as:

> participant-observational research based in online fieldwork. It uses computer-mediated communications as a source of data to arrive at the ethnographic understanding and representation of a cultural or communal phenomenon. … Netnography has its own uniquely adapted set of practices and procedures that set it apart from the conduct of face-to-face-ethnography. (p. 60)

Netnography and ethnography do share similar guiding assumptions, including that "personal engagement with the subject is the key to understanding a particular culture or social setting" (Jupp, 2006, p. 101). *The Sage Dictionary of Social Research Methods* (2006) adds that netnography "is a naturalistic and unobtrusive technique—a nearly unprecedented combination" and outlines its main differences from traditional, offline ethnography: the unique interface of the computer, which can be both constraining and liberating; the unprecedented access to particular cultures and behaviors; the automatic archiving of interactions and data; and the uncertainty of anonymity and identity (Jupp,

[16] The Digital Methods Initiative (2012) defines *post-demographics* as "the study of the personal data in social networking platforms. … With the 'post' prefix to demographics, the idea is to stand in contrast to how the study of demographics organizes groups, markets, and voters in society. The term post-demographics also invites new methods for the study of social networks, where of interest are not the traditional demographics of race, ethnicity, age, income, educational level or derivations thereof such as class, but rather tastes and other information supplied to make and maintain an online profile."

2006; Kozinets, 2010). Netnography shares with ethnography the assumption that "personal engagement with the subject is the key to understanding a particular culture or social setting," and document analysis and interviewing are common tools in both the netnographer's and ethnographer's repertoire (Jupp, 2006, p. 101). Finally, Kozinets (2010) cautions that netnographic techniques are particularly suited for studying "online communities," where phenomena is fundamentally tied to the online environment, but not as well suited for studying "communities online, " where online activities compliment an offline culture.

Though this study should not be characterized as a formal netnography, netnographic assumptions and techniques were used to capture and analyze data from the natively online community of Wikipedia. The document analysis and interviewing techniques outlined below accessed archival and elicited data, respectively. Fieldnotes, an additional form of data usually collected via netnography, were not systematically maintained, but a manner of fieldwork was conducted. The researcher regularly logged in to Internet Relay Chat (IRC) rooms associated with Wikipedia and the Wikipedia bot community, watching the public conversations and notifications in those forums. On Wikipedia, Talk pages related to policies, articles, and most notably, Bot Request for Approvals (BRFA), were also observed frequently. Finally, the researcher attended the 7th International Symposium on Wikis and Open Collaboration (WikiSym) in Mountain View, California, in October of 2011.[17] Though not documented in the research methodologies below, observations from this "virtual" fieldwork aided the researcher in

[17] While this was an offline meeting of Wikipedians and researchers, the data gathered could still be deemed "netnographic" as it served to compliment the larger dataset collected from the online community.

understanding certain nuances of the Wikipedia community that ultimately strengthened both the overall research design and the subsequent interpretation of data.

Document Analysis

Document analysis was the backbone methodology for this study, on which additional methods added flesh and muscle. Scott (1990) defines documents as "physically embodied texts, where the containment of the text is the primary purpose of the physical medium," (p. 13), but Lindlof and Taylor (2011) contextualize the importance of these objects in our contemporary life: "There is no getting around the fact that documents are deeply embedded in people's work and leisure worlds—as prompts to action, as informational resources, as aids in speech acts, as items of transaction, and so on" (p. 231). Indeed, Wikipedia—at its simplest, a collection of documents—is perhaps the quintessential exemplar of this definition in the context of digital life. A close examination of relevant documents, policies, and archives on the site, then, is vital to understanding the technical development of the medium and the social forces that drive this development.

Lindlof and Taylor (2011) highlight that the prime advantage of document analysis over other research techniques is the richness of information it can unlock: "that is, [documents] are richly infused with the history, idiomatic speech, and cultural logics of the people who made them" (p. 235). The value of each document, however, cannot be taken for granted, and Scott (1990) offers a concise rubric of four criteria on which to judge quality: authenticity, credibility, representativeness, and meaning. Although the examination of documents can be systematized and the data collected quantified, ultimately document analysis is an interpretive method, which its positivist critics deem a weakness. The application of Scott's (1990) rubric serves to solidify confidence in the data collected from documents and lead the researcher to imbedded truth claims in the source material.

For this study, document analysis both supported and complimented the other research methods: interviewing and social network analysis. As the first stage of the study, reviewing select primary sources on Wikipedia revealed key editors to target for the interview stage and provided guidance in constructing interview guides. Conversely, interview responses often identified additional documents and sources, both internal and external to the site, to review for further information and perspective. The use of these two methods together provided the ability to triangulate data and "bolster confidence in the objective reality of research findings" (Lindlof & Taylor, 2011, p. 274); the accounts and recollections of interviewees could be compared to the densely archived history of the site to verify claims. Wikipedia documents also supplied both raw data and contextual information for the social network analysis performed for the study.

Sample. Essentially, Wikipedia is a vast collection of three types of informational documents: content pages, discussion or Talk pages, and History pages. All content, whether an article, a proposal, a WikiProject, or a policy (to name a few) is associated with a Talk page where editors are encouraged to informally discuss, debate, and establish consensus before changes are made to actual content. History pages are automatically maintained by the MediaWiki software for both content pages and Talk pages; these History pages record the user, date, data size, and optional edit summary for all changes, as well as archive previous instances of the page in case newer revisions need to be reverted.[18]

Published under the Creative Commons Attribution-ShareAlike 3.0 License, all Wikipedia pages are free, open, and available to the Internet community. In addition, the persistent archiving of content, including all past revisions of pages, and the automatic logging of both page and user histories have created an exhaustive reserve of data on the site. Together, these attributes (open and exhaustive) offer the researcher numerous sampling options to match the nature of their specific research questions. Many computer science researchers of Wikipedia work with a complete version of the database (a data dump) to explore macro-level issues and trends (Kittur, Chi, & Suh, 2009; Ortega, 2009; Wilkerson & Huberman, 2007). Social scientists interested in more nuanced questions or more specific content and/or user groups have employed various sampling methods to meet their needs (Forte, Larco, & Bruckman, 2009; Konieczny, 2009; Oboler, Steinberg, & Stern, 2010). The sample for this study's document analysis primarily considered content and Talk pages, as these are the sites of qualitative data (History pages were later used for network analysis).

The selection of Wikipedia documents for analysis occurred in three phases. For the initial analysis, a purposive sample of English WP pages related to bots, bot policy, the Bots Approval Group (BAG), and the MediaWiki software were selected (see Appendix XXX for a full list of pages). These choices were theoretically grounded in the framework of the study and its interest in understanding how human actors and technological tools create a network or system of social and vocational meanings. Most of these documents (with the exceptions of current, *live* content and very recent

[18] An in-depth analysis of wiki pages is offered in Chapters IV and V.

discussions) exist in *archived* form on the site, meaning they are fully preserved and maintained in their entirety, but editors are alerted not to change the contents of the pages. In total, this sample of documents represented approximately 900 PDF pages of text.

From references to policies, events, incidents, anecdotes, people, and other sources within the initial sample, a snowball sample of additional documents was identified. As these documents were examined, further sources were identified and the process was repeated recursively until a sufficient number of documents were obtained.[19] Among the most significant of these additional documents were relevant Bot Request for Approval (BRFA) pages, present and past issues of Wikipedia's community-written newspaper *The Signpost*, and discussions of technical issues and policies at Wikipedia's technical forum, the Village Pump.

A third sample of documents was identified from suggestions and references made by interview subjects. As discussed later in this chapter, interviewing in the online environment offers the ability for the researcher and subject to exchange direct links to documents and websites during the conversation, and this was an active part of most interviews for this study. Some of the sources that emerged from these suggestions included bot policies, documentation, and discussions at the Wikimedia Foundation and Wikimedia Commons, resources and information on WMF data centers and the Wikimedia Toolserver, pages on interwiki bots, and complementary websites like Botwiki.

[19] Please note: Wikipedia contains a virtually endless "paper trail" with information spread out across many different pages; while the size of a sample of documents should depend ultimately on the research questions and the point of information saturation, the realistic limits of the researcher's time and energy also play a role in determining the sample size.

Finally, secondary sources outside the immediate purview of Wikipedia were consulted to provide context and perspective from outside the organization. Articles in newspapers and periodicals were located using the Factiva and Lexis-Nexis Academic online databases, and additional Web sources (blog posts, websites, interviews, commentaries, etc.) were found through Google searches.

Interviews

To elucidate the social and cultural meanings, understandings, and experiences that underlie the technical development of Wikipedia, qualitative interviewing was employed in this research as a second major data collection method. Rubin and Rubin (2005) argue that "qualitative interviewing projects are especially good at describing social and political processes, that is, how and why things change" (p. 2). Lindlof and Taylor (2011) add that interviewing is "remarkably adaptable to varied circumstances and settings," including the online environment (p. 171), while Kozinets (2010) maintains that methodologically, interviewing and netnography are "virtually inseparable" (p. 46).

The interviewing for this research was theoretically grounded in interpretive constructionism, which holds that meanings and definitions of phenomena are created, maintained, and revised by actors in a cultural arena (Rubin & Rubin, 2005). The work of the constructionist researcher is:

> to elicit the interviewee's views of their worlds, their work, and the events they have experienced or observed. Rather than looking for the average and ignoring the specific, as positivists often do, interpretive constructionists look for the specific and detailed and try to build an understanding based on those specifics. (Rubin & Rubin, 2005, p. 28)

This is not to deny that quantitative data can be collected and statistical measures performed on interview data, but only that the meaning and significance of all types of research collected must be interpreted from the cultural lenses of the participants. "Everything that people say in an interview," argue Lindlof and Taylor (2011), "issue from a perspective that is uniquely their own. … Whether they realize it or not, they are the authors of the stories they tell" (p. 173).

Interpretive constructionism is a particularly telling framework in which to study the Wikipedia community, as maintaining a neutral point of view (NPOV) is a cornerstone community value for the content that editors produce. Interviews rooted in the qualitative tradition, then, offer a method for getting at the underlying conflicts and collaborations that find little acceptance in the main article space of the site. The method used in this research closely aligns with what Rubin and Rubin (2005) term "responsive interviewing," where the goal is to generate "depth of understanding, rather than breadth" while remaining flexible to new paths emerging during the interview (p. 30).

Sample. Though the qualitative inquiry of this research project took many cues from the ethnographic—and newer netnographic—traditions, interviews were not quite as free-flowing as ethnographic interviews often are, and interviewees were largely targeted in advance. The bulk of interviewees straddled Lindlof and Taylor's (2011) categories of "informant" and "respondent." Informants "inform the researcher about the scene—the scene's history, customs, and rituals; the local 'lingo'; the identities and actions of the key players; and so forth," while respondents speak mostly about themselves—their attitudes, opinions, and experiences (p. 177, 179). Other common forms of interviews (narrative, focus groups, etc.) were not used in this study.

A comprehensive sample of interviewees for the project was gathered in two ways. First, a sample of interviewees was drawn from all English WP editors involved in creating bots and assisted editing tools for the site. This includes editors whose bots currently have approval, or a *bot flag* (n = 699) or were currently undergoing the Bots Request for Approval (BRFA) process as of January 1, 2011 (n = 13). The BRFA process is facilitated by the Bots Approval Group (BAG), a board largely composed of Wikipedia Administrators (Admins) and fellow bot operators. As such, BAG members were also contacted for interviews; the group includes 16 active members, 28 inactive members, and 18 former members.

Three outreach methods were used to contact this sample of interviewees, each tailored to the particular audience. First, a public invitation to participate in the study was posted to a number of Wikipedia pages related to the development of bots, technical tools, and the technical infrastructure of the site, requesting interested editors contact the researcher directly via email or Talk pages (Appendix XXX). Second, targeted subjects were contacted directly via email or their Talk pages; this included all past and present members of the BAG, the operators of the top 100 bots by edit count, and editors who actively participated in the creation of and discussion around Bot Policy (as found from reviewing the Bot Policy archived discussion pages). Third, at the conclusion of each interview, subjects were asked to recommend the study to fellow Wikipedia developers and collaborators, thus employing a snowball method of recruitment.

A second sample of interviewees was drawn from the staff and volunteers at the Wikimedia Foundation involved in the technical development of the organization's projects or in a position to comment on the governance and development of technical issues, as well as developers of the MediaWiki platform (n = 68). Direct outreach, as described above, was made to WMF staff and volunteers in positions relevant to this study. Demographic and descriptive data on the final interview sample is presented in Figure 2.

Interviewing Methods. The majority (96%) of interviews for this study were conducted online via instant chat (36%), email (53%), or video chat (7%). As Wikipedia is a virtual, dynamic, and internet-mediated project where community performance happens online, these formats were both intuitive and preferred for the bulk of interviewees.

The methodological literature on computer-mediated interviewing has greatly expanded in recent years, documenting both the strengths and weaknesses of these interactions (Markham & Baym, 2009). Both live, synchronous chat and asynchronous email offer some common benefits to the researcher and his subjects. Kozinets (2010) highlights the inherent ability of computer-mediated interviews to be automatically archived (transcribed and saved), which means the "researcher can be freed from routine note-taking or transcription concerns to concentrate fully upon the body of the interview" (p. 111). Online interviews also lower the costs of collecting data (no traveling expenses, telephone charges, transcription fees) and can offer increased access to specialized populations (Fontana & Frey, 2005; Jupp, 2006). For this study, a major benefit of the online environment was the ability for interviewees to present hyperlinks to relevant examples and archived discussions; this eliminated a constant need to fully summarize previously documented events and allowed subjects to focus more on their experiences and opinions.

Of course, to take advantage of computer-mediated interviewing techniques, weaknesses of the medium must be negotiated. Most notable of these weaknesses is the

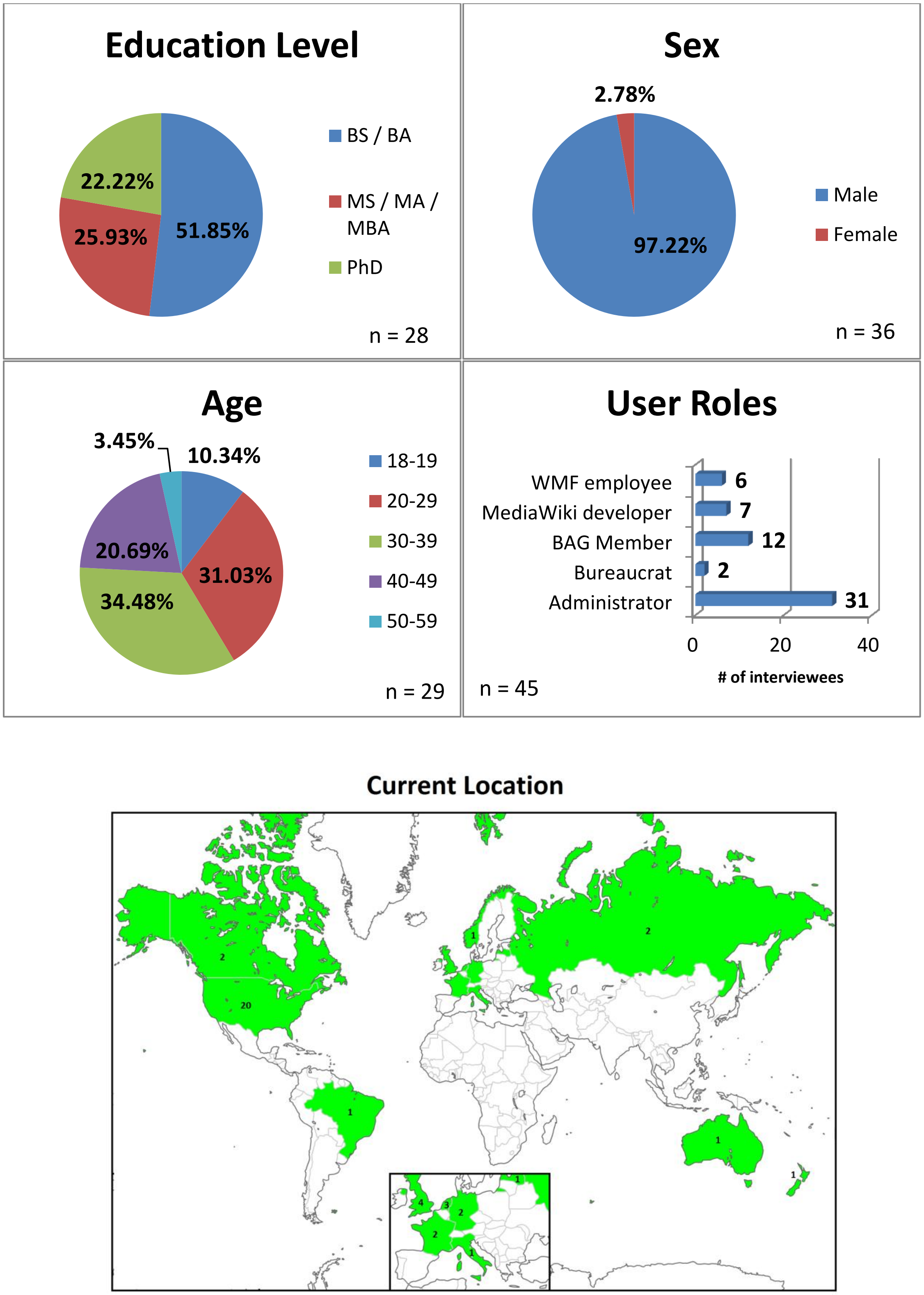

Figure 2. Descriptive statistics for the interview sample. Note: Interviewees were given the option of disclosing individual demographic data, thus producing different n values for each category.

absence of nonverbal cues and behaviors available to the researcher, though online-specific conventions (discussed below) have emerged to somewhat gauge the feelings and emotions of the subject. The lack of physical familiarity, which can generally be gained in face-to-face interviews, poses a challenge to establishing a trust relationship between interviewer and interviewee (Fontana & Frey, 2005) and must be compensated for in the online presentation of self; as Hine (2000) points out, care must be taken with all online communication to subjects, because "self-presentation is crucial in forming relationships with potential informants in online settings" (p. 74). Other disadvantages to online interviewing include the difficulty in confirming the identity of the subject (i.e. the possibility of deception is higher) and the difficulty in maintaining anonymity in an environment that automatically documents all communication (Fontana & Frey, 2005). For this study, the former concern was mitigated by comparing interview data with documented data on Wikipedia about the user; the latter concern was avoided through the informed consent process.

Online Chat. Berg (2007) argues that synchronous online environments like private chat rooms provides "the researcher and respondent an experience similar to face-to-face interaction insofar as they provide a mechanism for a back-and-forth exchange of questions and answers in what is almost real time" (p. 112). The implications of this include the ability for the researcher to ask immediate follow-up and/or clarification questions, or to deviate from the structured interview questions to follow new threads of conversation that emerge in the dialogue. In addition, although traditional non-verbal cues are not present for the researcher to "read," online chat offers its own set of conventions to gauge the state of the interview subject, including the length of pauses, the use of emoticons, and the informal grammar of chat rapport (Lindlof & Taylor, 2011).

For this study, three internet chat clients were used to conduct synchronous online interviews: Internet Relay Chat (IRC) via the Chatzilla plug-in for Firefox, Gchat, and Skype. These clients are roughly equivalent in terms of usability, though IRC does not cue when the chat counterpart is typing; this prompts a bit more dialogue to distinguish whether the subject is thinking, typing out an answer, or done with her response to the question. Interview appointments were set up about one week in advance via email or Talk pages, and each interview subject chose the chat client they preferred for the interview. The researcher followed a semi-structured question guide that was tailored to each subject (as discussed later), but asked real-time follow-up questions as they presented themselves. Interviews normally lasted between 60 and 90 minutes, and additional follow-up or clarification questions were often sent via email over the next week as the researcher analyzed the interview data. Of the 16 interviews conducted via online chat, seven utilized IRC, six utilized Gchat, and three utilized Skype.

***Skype*.** Skype video chat was also used to conduct synchronous interviews for the study. Video chat offers some of the benefits of traditional face-to-face interviews (the ability to establish rapport through eye contact and to read nonverbal language; typing is unnecessary, making the experience a bit easier for the subject) while maintaining the benefits of online chat (passing links back and forth; facilitating real-time communication with subjects from around the world). The major drawback of video chat is that the conversation is not automatically archived, so interviews needed to be audio recorded and later transcribed. Other disadvantages sometimes associated with video chat include slow

Internet connection speeds or dropped connections, calling fees or other charges, and reduced anonymity, but none of these proved problematic for this study.

Email. Asynchronous environments like email offer their own advantages. Berg (2007) highlights that "the use of email has become a common and comfortable activity. Transferring this comfort to the interview situation, then, can similarly provide a benefit for qualitative interviewing" (p. 113). Surveying the recent research, Lindlof and Taylor (2007) found that people often enjoy the flexibility of an email exchange, as it offers subjects both control over the time spent responding (which can be parceled out to fit their schedule) and the ability to contemplate their responses before submitting them, thus creating "a more fully reflexive interview" (p. 191). Kozinets (2010) adds that the "persistent conversation" offered through email exchange "can lead to revelatory personal and emotional discovery" (p. 112).

Interview subjects for this study were emailed the individually customized question guide within one week of initially contacting the researcher. Questions were sent as a Word .doc attachment, and the following message was conveying in the body of the email:

> These questions are designed to get at your experiences with and opinions of the bot community and technical infrastructure on Wikipedia, as well as some feelings on Wikipedia as a whole. Having said that, these questions are not hard and fast...they are a guide. Please respond in as much detail as you find necessary, but feel free to discuss tangential issues or examples that come to mind. Of course, you are free to not answer or skip any questions that you like. If a question is confusing and you'd like more clarification, please let me know.

Subjects were asked to respond within one month. Once a response was received, follow-up or clarification questions were sent over the next week, as needed.

Interview Questions. Semi-structured interviews were conducted utilizing an interview guide built around key topics, ideas, and themes that spoke to the research questions. Interview guides offer an "informal, flexible approach" that can be tailored to different groups of participants and allows the researcher the freedom to add questions, drop questions, or pursue tangential topics (Lindlof & Taylor, 2011, p. 200). Although all interviewees for this study had a stake in the technical side of Wikipedia, their roles and activities, group memberships and allegiances varied, so the adaptability of semi-structured interviewing was optimal. Interview questions were developed based on previous Wikipedia research and subjects that emerged in the initial stage of document analysis. Within the interview guide, three types of questions were constructed: main questions, follow-up questions, and probes. As Rubin & Rubin (2005) explain, "The main questions help you make sure you are answering your research puzzle; the follow-up questions and probes ensure that you get depth, detail, vividness, richness, and nuance" (p. 129).

The interview guide contained a consistent framework of questions for all interviews, which aided in the analysis and coding of data, but numerous questions were personalized for each interview subject to establish rapport, build confidence and credibility in the researcher, and maintain flow during the dialogue (see Appendix XXX for the complete interview guide). These small details—for example, using the specific name of an interviewee's bot, or referencing a piece of information on his or her user page—showed the subject that the researcher "did his homework" and was invested in that specific interview. In all formats of interviewing, this technique helped to build a relationship that aided in later inquiries for follow-up or clarification information.

Social Network Analysis

Broadly defined, social network analysis (SNA) is the "analysis of any pattern of relations between individuals, groups, institutions or any other social collectivities" (Jupp, 2006). A form of structural analysis whose basic unit of study is the relationship, the method's origins lie in sociology and graph theory, though it is now employed in a wide range of fields from computer science to business to communication studies. This adaptability emerges from the fact that SNA can be used both quantitatively and qualitatively to examine and visualize multifaceted data on the relationships between social actors (Carrington, Scott, & Wasserman, 2005; Scott, 1991; Vego-Redondo, 2007; Watts, 2003). Existing research on Wikipedia has employed social network analysis as a method for investigating how editors work together to produce content (Jesus, 2010; Kane, 2009; Livingstone, 2010), as well as for visualizing the development and evolution of content (Kimmerle, Moskaliuk, Harrer, & Cress, 2010).

Population and Boundaries. Although entire data dumps of the English WP, comprising information on over 3.8 million articles and over 16 million unique user accounts, are freely available, such a large dataset was too unwieldy for this study. More importantly, though, a complete data dump was unnecessary to effectively address this study's research questions, which revolve around understanding the relationships between human users, bots, and the content that they produce. Rather than using traditional forms of statistical sampling, which is problematic for network methods (Rothenberg, 1995), social network analysts study populations of actors that fall within certain boundary parameters. These boundaries can be of two main types: those created by the actors themselves, and those defined by the researcher based on relevant criteria

(Hanneman & Riddle, 2005). The networks explored for this study, specific groups of Featured Articles on the English WP, are bounded in both ways.

Featured Articles are "the best articles in Wikipedia, as determined by Wikipedia's editors" (Wikipedia, 2012y), and as Ayers, Matthews, and Yates (2008) explain:

> … go through a more formal community peer review process [than most articles], typically with several different editors participating as reviewers. Reviewer comments are likely to be detailed and extensive, ranging from minor issues, such as formatting, to major issues, such as unclear writing or missing references. (p. 228)

At the time of this study, over 3,550 articles were designated as Featured Articles on the English WP, and a number of studies on Wikipedia content have considered Featured Articles as units of analysis (Goldspink, 2010; Kane, Majchrak, Johnson, & Chenisern, 2009; Viégas, Wattenberg, & McKeon, 2007); however, none have used these articles to explore editor activity. Featured Articles represent a self-bounded population of editing activity, while at the same time, a researcher-defined population of articles with sufficient revision histories to explore the relationships between human editors, bots, and quality content.

Data Collection and Analysis. Data collection was facilitated by the use of two online tools: the Contributors report script run from Wikimedia's Toolserver, and a local collection script developed for this project. The Wikimedia Toolserver is a hardware cluster operated by the German Wikipedia that hosts "various software tools written and used by Wikipedia editors" (Wikimedia Meta-Wiki, 2012f). The Contributors script, written by User Duesentrieb, aggregates and reports data on all contributions to a particular queried Wikipedia article; this data includes each contributor's username, number of edits (additions and reversions), date of first edit, and date of most recent edit. Data may then be filtered or sorted by a number of different criteria.

To collect a comprehensive dataset that includes the contribution histories of all Featured Articles, a software script, Scwape, was written by a third party using the REBOL programming language and run from the researcher's local computer. Scwape would systematically query the Contributors tool for all articles listed on the Wikipedia:Featured Articles page and amass the returned data into one tab-separated (TSV) data file. Full execution of the script takes approximately eight hours, and Scwape was run on January 8, 2012 to collect data for this study. The size of the resulting TSV file was 105.8MB.

The complete Featured Articles data was imported to Microsoft Excel and parsed according to the predefined content area categories that organize the Featured Articles page. Two categories—Food and Drink, and Philosophy and Psychology—were selected for further exploration based on their rich but manageable size. The articles in each category exhibit a sizable variance in number of contributors and number of edits, and yet all have been selected by the Wikipedia community as representative of the best content on the site. Table 1 presents descriptive statistics on all articles in these samples.

Data Analysis

Lindlof and Taylor (2011) write that one of the biggest challenges with qualitative research is "just coping with all of the data that must be 'processed,' understood, and turned into a useful contribution" (p. 241). To meet this challenge, the researcher must

Table 1

Descriptive Statistics for Articles in Network Samples

Food and Drink	Total Edits	Total Editors	Date created	Date promoted to Featured Article
Durian	2627	936	10/25/2001	3/26/2007
Ina Garten	1339	515	5/20/2004	4/8/2006
Gumbo	865	465	8/5/2003	7/15/2011
History of saffron	467	238	12/24/2005	1/10/2006
Malagasy cuisine	858	184	7/23/2007	7/19/2011
Maple syrup	2265	1118	9/9/2001	8/26/2011
Maraba coffee	372	138	6/8/2006	10/6/2006
Medieval cuisine	1509	627	9/17/2006	3/9/2007
Odwalla	577	194	8/20/2004	9/23/2008
Saffron	2152	808	5/24/2002	12/27/2005
Thomcord	92	24	1/2/2011	2/8/2011
Trade and use of saffron	411	160	1/22/2006	2/4/2006
George Washington (inventor)	659	333	3/27/2007	3/31/2007
Total Edits	14193			
Total Unique Editors		5052		

Philosophy and Psychology	Total Edits	Total Editors	Date created	Date promoted to Featured Article
Attachment theory	3424	566	8/8/2004	11/30/2009
Conatus	763	139	8/5/2006	5/5/2007
Confirmation bias	1087	347	6/24/2002	7/6/2010
Eric A. Havelock	358	166	2/13/2006	3/8/2006
Free Will	2869	1107	4/6/2002	1/19/2004
Philosophy of mind	1895	756	6/4/2001	3/27/2006
Hilary Putman	1077	400	8/23/2002	8/16/2006
Reactive attachment disorder	2250	296	6/21/2004	3/4/2008
Transhumanism	6744	1084	7/4/2001	5/14/2006
A Vindication of the Rights of Women	829	288	12/8/2001	3/20/2007
Bernad Williams	725	294	3/25/2004	12/7/2004
Total Edits	22021			
Total Unique Editors		4749		

have a plan—a thoughtful and purposeful way to sift through the data in a systematic, organized fashion that will reveal patterns, themes, and meanings to address the study's research questions. In a sense, the researcher performs a content analysis of the data, ever mindful that the process involves a keen attention to detail, careful interpretation, and an awareness of one's own biases and personal position relative to the subject matter.

The qualitative data for this project, primarily collected through document analysis and interviewing, was analyzed using a standard, established process as outlined by Lindlof and Taylor (2011), Berg (2007), and others. This process largely revolves around three major steps: examining the data, reducing and sorting the data into categories and codes, and interpreting the categorized data. Examining the data requires numerous close readings for both denotative and connotative meanings, which involves an element of interpretation even at this first step. Categorization then takes place, where units of data are sorted "with respect to properties that they have in common," often using semantic codes to identify individual units (Lindlof & Taylor, 2011, p. 246). In the present research, conceptual categories emerged inductively after codes were assigned and data was grouped together. This formation of categories established the foundation for the in-depth interpretation of data, where the theories and concepts that guided the initial creation of data-gathering instruments are reintroduced to create and validate claims (Lindlof & Taylor, 2011). Though these steps may flow sequentially, they often overlap in a continual process of refinement, with the researcher going back and forth to reread, recategorize, and ultimately further clarify the meanings embedded in the data. As such, qualitative data analysis is a heavily recursive process.

To begin data analysis for this project, all data was pasted into Microsoft OneNote, an electronic information organization and note-taking tool. Data included all notes from the document analysis, selected primary documents, all complete interview transcripts, and any notes or memos written during the data collection stage of research. This data was originally aggregated in three tabbed pages of OneNote ("Documents," "Interviews," and "Notes"), with subpages holding individual sets of data (a single interview transcript, for example). Interview data was examined first and sorted in two ways. As a semi-structured interview guide was employed during data collection, responses were split onto separate pages based on common questions. Additional responses that resulted from follow-up questions, points of clarification, or tangents in the dialogue were moved to supplementary pages built around emergent categories (for example, "MediaWiki code development" or "Bureaucracy"). Data from the document analysis was then categorized in a similar manner, followed by the categorization of personal research notes. The resulting organization grouped all data into two main pages, "Interview Q Responses" and "Emergent Categories," with 24 and 41 subpages of sorted data, respectively.

Next, all subpages were grouped into more general meta-categories. These meta-categories emerged from examining the dominant themes of the subcategories, but they were also heavily informed by the researcher's experiential knowledge gained from interacting with the Wikipedia community and reviewing pertinent literature, as well as from an awareness of the project's main research questions. The following 13 meta-categories were developed: social, tasks and permissions, collaboration and conflict, technical performance, MediaWiki, Wikimedia Foundation, bots (general), bot policy, Bot Approvals Group, global bots, community reaction to bots, semi-automated tools, and collective intelligence. (See Appendix XXX for a list of subcategories within each meta-category.) These meta-categories were then printed; hardcopies of the categorized data were used for the remainder of the analysis to make final interpretive notes, to identify revealing and exemplary quotations, and to reference during the writing process.

Network data was handled and analyzed using two additional software programs. Data collected by the Scwape software script was opened in Microsoft Excel, where it was cleaned, parsed, sorted, and finally converted into a 2-mode matrix (editors x articles). UCINET was then used to calculate network measures and visualize network matrices, utilizing the NetDraw subroutine for the latter task. Interpretations of these networks were then made in the context of the study's main research questions and the interpretations of the other methods.

Limitations

Many of the methodological limitations of this research have been mentioned in the preceding descriptions of the study's planning and data collection, but there are other limitations that are important to acknowledge. In an effort to reach a particular population of Wikipedia contributors (bot operators) who are conceptually situated in a unique position for understanding the site as a sociotechnical system, other stakeholder groups were passed over. Though a limited number of Wikimedia Foundation employees and MediaWiki developers were interviewed, a much broader sample of these actors and contributors would be necessary to make strong, detailed statements about their particular roles in the network of collaboration that produces the site. In addition, the attitudes and opinions of general contributors to Wikipedia were not solicited, nor were general readers of the site. Thus, conclusions made by the researcher about how the system works were largely conditioned by the perspective of the interview sample, a group of highly technology-literate, invested contributors.

Another limitation of this research was the lack of attention given to individual content on Wikipedia. The focus here was on contributors (both human and technical actors) to the site, with the assumption that attitudes, opinions, and actions largely open a window to observe the potential for sociotechnical collaboration and collective intelligence. However, the evaluation of content likely provides another avenue for investigating these phenomena, with Niederer and van Dijck (2010) suggesting the "technicity of content," explored through natively digital methods of content analysis, offers another powerful perspective for understanding Wikipedia as a sociotechnical system. Though some of the work behind specific content on the site—data on edit histories—is used in the exploratory network analysis, this study does not consider content per se, and thus makes few assertions about the semantic or discursive nature of Wikipedia articles.

Finally, as Rubin and Rubin (2005), among others, emphasize, neutrality in qualitative research is impossible, and the researcher's positionality and personal biases undoubtedly influence his interpretation of data. As is much of the Internet community, I am a daily user of Wikipedia as a reference tool, both for personal and academic purposes. I began researching Wikipedia for a separate project in early 2010 and have made a small amount of contributions to articles since learning how the system works. I have also attended an international conference largely devoted to Wikipedia (WikiSym '11), met numerous contributors, developers, and Wikimedia Foundation employees, and informally chatted online with an array of Wikipedians. Overall, I highly respect this open-source community as well as the mission of its governing organization. As a researcher, these experiences and attitudes, as well as other personal theoretical orientations and viewpoints, are important to acknowledge, as they could affect the interpretations and conclusions of the study. To mitigate against the possibly harmful effects of personal bias, data triangulation was employed whenever possible to verify facts, flesh out interpretations, and legitimize claims.

CHAPTER IV

THE BIRTH AND DEVELOPMENT OF WIKIPEDIA'S TECHNICAL INFRASTRUCTURE

The idea for a universal encyclopedia of all human knowledge has been around for at least two millennia. The major attempts to make this idea a reality share a common link beyond the obviously ambitious task at their core, though. Each attempt has been thoroughly conditioned by the social environment of its creators, reflecting the cultural epistemologies and perspectives on knowing the world that were dominant in that time and place, from the Roman Empire, to the Christian Middle Ages, to the Enlightenment, to the Victorian era. To understand, appreciate, and critique such a work, then, requires careful consideration of the social, technical, and economic contexts in which it was produced.

The latest incarnation of the universal encyclopedia, Wikipedia, is no different in this respect, as it is a product of its times: the Computer Age. Stretching as far back as the 1940s, the Computer Age has unfolded in many unpredictable ways, however, with the emergence of the Internet and the open-source software movement in the late 1960s and early 1970s fighting against the dominant 20th century paradigm of capitalism-driven society and technology. In the wake of these developments, we live in a digital world of enhanced access and communication, where a project to build an encyclopedia becomes an experiment in community, cooperation, and code unlike any before.

Wikipedia exists at the intersection of dueling ideologies—postindustrial capitalism and the open-source technology movement—each of which has influenced the development of its technical infrastructure, from the MediaWiki software at its core, to its hardware inventory and data centers, to its technical staff and developer community. The site's history, including its dot-com era genesis, adaptation of wiki technology, reliance on volunteer labor, and formal organization as a non-profit, reflects this dual nature. So to begin understanding Wikipedia as a sociotechnical system, we need to trace how these forces have historically conditioned the project, and how they continue to shape its development.

This chapter does not aim to retell the well-chronicled events that led to the birth of Wikipedia and its subsequent growth to a top-10 global website and cultural icon. Instead, this chapter offers a more focused history of the technical development of Wikipedia to inform the project's overall argument that the site flourishes because of its sociotechnical nature. What actors and forces have shaped the technical community that maintains and grows Wikipedia? How and why has the computer infrastructure—the servers and software that form the backbone of the system—developed as it has? How do material and immaterial factors influence Wikipedia's technology? These questions will be addressed in three parts. First, the context for Wikipedia's genesis, both as an encyclopedia and a website, is provided by presenting key concepts and developments from the computer and information revolutions. Then, using the structure of a Wikipedia page as a guide, key events from the history of the site, including the decision to use a wiki framework, are explored to understand the technical direction Wikipedia has taken. Finally, a political economic analysis explores how the site leverages its unique position in the Internet ecosystem to build and maintain its infrastructure.

In many ways, this chapter provides context for the project by investigating the sociotechnical context of Wikipedia's development. Research in the discipline of historical technology studies often takes either an internalist or externalist (contextualist) approach. The former "maintains that we can understand the development of a technology only if we start with an understanding of the technology in all its minute details," while the latter "claims that the economic, social, political, and scientific context of a technology is as important to its development as are its technical design characteristics" (Bijker, 1995a, p. 9-10). As Bijker (1995a) does in his analysis of the bicycle and the fluorescent bulb, the course for this research is to "step outside the distinction between technology and its context" and reject determinism by understanding the nuanced relationships between the historical forces, social actors, and technological designs involved in the creation and evolution of digital networks, wiki technology, and Wikipedia itself.

A (Very) Brief History of the Encyclopedia

Despite his eventual alienation and exit from the site the following year, Wikipedia cofounder Larry Sanger (2001) strongly believed the ways in which knowledge is collected and distributed were shifting quickly in the Internet era. He prophetically proclaimed: "The grandest days of free content have not yet begun. Britannica and other proprietary encyclopedias will be hopelessly obsolete within ten years—small, out-of-date, and generally irrelevant—by comparison with Nupedia, Wikipedia, and the many other non-proprietary reference works that are being and will be developed" (Sanger, 2001b).

Sanger was off by about a year. On March 13, 2012, Encyclopaedia Britannica Inc. announced it would discontinue the print edition of its flagship work after 244 years, instead focusing on its online version (Bosman, 2012). Eulogies and laments immediately hit the Web, but some, like opinion writer Jim Sollisch (2012), argue for a positive interpretation that largely mirrors Sanger's earlier point:

> Let us not praise gold-leafed-leather-bound knowledge. Rather, let us trumpet its passing. Let's celebrate the fact that information now roams free, great herds of it. It lives and breathes, loosed from cages where it was allowed to reproduce only once a year, edition by edition.

In nearly all reports, Wikipedia is cited as the death knell to Britannica's hold on authoritative encyclopedic knowledge, with the online experiment in collaborative authorship and free content now far surpassing its forbearer in depth, breadth, and most significantly, readership.

Fowler (1996) writes, "Every age has its particular encyclopedia" (p. 7). For the contemporary digital age, that work is Wikipedia, a dynamic site that both follows a long tradition of encyclopedic efforts and breaks away from tradition in new ways. Other Wikipedia researchers have traced in detail the historical "pursuit of the universal encyclopedia," including Wikipedia's place in this history (O'Sullivan, 2011; Reagle, 2010). For this project, only a brief recap of the philosophical and material bases of such encyclopedic works is necessary to understand the emergence of Wikipedia.

Though the term as we know it today was popularized during the Renaissance, the etymology of "encyclopaedia" is from the Greek *enkyklos paideia*, meaning "the circle of subjects" (Fowler, 1996). Dating from Plato's Academy, this original meaning signaled a course of instruction to educate a young mind, and unlike the later definition meaning "a literary work of reference," an encyclopaedia was intended to be used as a linear curriculum of education (Collison, 1964). There is little evidence that the ancient Greeks attempted a comprehensive written encyclopaedia, and the Roman Pliny the Elder is largely acknowledged as the first compiler of a reference encyclopaedia[20] (Collison, 1964; O'Sullivan, 2011). From the *Naturalis Histoia* of Pliny through modern incarnations in print and online, encyclopaedias have come to be associated with comprehensive reference works, or repositories of knowledge, that are read and consulted in a non-linear fashion; the user reads bits and passages as needed, but not cover-to-cover.

The history of the encyclopedia has been shaped by numerous thinkers, each putting forth ideas on the nature of knowledge influenced by the period in which they emerged. *Speculum Maius*, the Middle Ages work of Vincent de Beauvais, was a Christian-influenced work meant to mirror the natural, unchanging order of the world set by God, while Francis Bacon, writing during the early scientific revolution of the 17th century, argued that knowledge requires progressive interpretation (Collison, 1964). The English *Cyclopaedia* and the French *Encyclopédie* emerged during the Enlightenment as radical works reflecting the new age of science, reason, and open information, while the *Encyclopædia Britannica*, largely conservative in its representations of both the natural and human spheres of knowledge, became "a symbol of authority and permanence in the middle-class Anglophone household" (O'Sullivan, 2011, p. 44). Though these multi-volume works are impressive in their size and coverage, their editors by this time had largely abandoned visions of chronicling the sum of all knowledge, instead offering a collection of the most important and necessary information.

Collison (1964) highlights the political and economic influences on encyclopedic content and distribution in the 19th and 20th centuries. Publications that attempted to aggregate both general and specialized knowledge into one work were threatening to

[20] The first computer server solely devoted to Wikipedia was named "pliny" in reference to Pliny the Elder.

other publishers and booksellers, who disparaged and attempted to discredit encyclopedias. At the same time, encyclopedias were costly works for families to keep on hand as reference when needed, so publishers would bundle volumes with other more popular works in order to make sales. The profit motive caused editors to remain conservative in their representations of the world, especially in articles about living persons.[21]

Reagle (2010) outlines many of the encyclopedias of the digital age and the technological and organizational ideas that enabled them. Vannevar Bush's memex, Paul Otlet's universal bibliographic repertory, and H. G. Wells' "world brain" all offered visions on how knowledge and information could be classified, compartmentalized, linked, universalized, and collaboratively produced. Project Xanadu, originally developed in 1960, was the first hypertext project to digitally link information from different computer files and provided both inspiration and challenge to the later design of the World Wide Web. Project Gutenberg, anticipating the personal computer revolution, promoted free access to knowledge in the public domain, and beginning in 1971 with the U. S. Declaration of Independence, the project archived literary and cultural works on its server for remote online access. Finally, with access to the Internet via home computers vastly expanding in the 1990s, discussions of a free and open encyclopedia emerged on mailing lists and bulletin boards, prompting inspired but unfinished projects like Interpedia, the Distributed Encyclopedia, and GNUpedia.

Fowler (1996) anticipated that online technology would move the world farther from the comprehensive chronicle of all human knowledge Pliny imaged: "The Internet

[21] Wikipedia also considers information on living persons particularly sensitive, offering detailed guidelines on how such biographies should be written and maintained (Wikipedia, 2012r).

gives the impression that knowledge is infinite and thus uncontrollable. The very idea of the sum of all knowledge seems to be disappearing" (p. 25). But the emergence of Wikipedia has challenged this notion, and by mobilizing a global, computer-literate user base, has made access to the world's knowledge the centerpiece of its mission.

Information

As the history of the encyclopedia indicates, the aggregation and assessment of knowledge has been a long-standing human pursuit. Wikipedia is the latest attempt in this pursuit, and perhaps the most successful attempt at drawing together "the sum of all human knowledge." But in the encyclopedia's functional turn from "the circle of knowledge" to "a literary work of reference" lay a potential gap between representations of knowledge and those of mere information. Indeed, we currently live in what has been popularly deemed the *Information Age* (Castells, 1996; Gleick, 2011), rather than something more akin to the "Age of Knowledge." The Information Age is often used synonymously with the *Computer Age*, as the lifeblood of ICTs and computer technology—from Shannon's (1948) theory onward—has been information as data, functionally abstracted from the context necessary to transform it into knowledge. It is important, then, to understand Wikipedia as an encyclopedia both from the tradition of its genre and the tradition of its medium.

The Need for Information

The roots of our current Information Age can be found in many of the economic and social changes introduced during the Industrial Revolution of the 18th and 19th centuries (Beniger, 1989). As enclosure movements privatized lands and precipitated urbanization, systems of mechanized production and factory labor emerged (Marx, 1867).

With the increase in productive capacity came a need for information management and control, and this need, in turn, drove many developments; socially, organizational bureaucracy and systems of control (Fordism, Taylorism) emerged to manage labor and maximize profits, while applied statistical and accounting methods were developed to track information and improve efficiency (Beniger, 1989; Braverman, 1974). Technologically, machines were invented to manage and manipulate information, from Babbage's Analytical Engine[22] to Herman Hollerith's electric tabulating system for the 1890 U.S. Census. Hollerith's machine, which used punched cards to feed data into a machine that could compute detailed statistics, was a public success, appearing on the cover of *Scientific American* and prompting him to found the Tabulating Machine Company[23] in 1895 (Computer History Museum, 2012).

Castells (2004) also acknowledges the significance of railroads, ocean liners, and the telegraph to the development of the current Information Age. Though these transportation and communication technologies were still hierarchical in many ways, they proved the importance of information and communication to the success of large-scale networks and offered models for information infrastructures, which Edwards (2003) defines as "ways to handle the functional problems of information storage, transfer, access, and retrieval (p. 207). Despite its theoretical separation of information from meaning, the emergence of information theory in the middle of the 20th century—the basis of modern information infrastructures—was itself a product of socioeconomic

[22] Described in Chapter II.

[23] The company later merged with two other firms to establish the Computing Tabulating Recording Company, renamed International Business Machines (IBM) in 1924. IBM would become one of the largest and most profitable U.S. firms of the 20th century, solidifying the importance of information technology in the growing economy.

forces; Bell Labs was a center for government-funded research during and after World War II. As discussed in the next section, the post-war climate in the West played a major role in the development of Internet technology.

Evolution of the Internet

Edwards (2010) argues:

> Like all wonders, and precisely because of its pervasive presence in every corner of modern life, the Internet's history lies shrouded in myth. The truth-value of such founding myths rarely matters as much as their dramatic and moral qualities, and the Internet is no exception. (p. 141)

Internet myths are largely fueled by technological determinism, or the idea that a series of early technological developments inevitably lead to further developments, fostering a spirit of invention and progress independent of social, political, and economic influences. Today, the taken-for-granted and ubiquitous nature of the Internet often seems to push these influences even further into the background.[24] As many Internet and media scholars have revealed, though, the Internet has been wrapped up in sociopolitical and socioeconomic issues from its genesis, with notions of information at the heart of the story (Abbate, 1999; Edwards, 2010; Gleick, 2011; Zittrain, 2008).

Shannon's (1948) information theory and the work of engineers and scientists at the Bell Laboratories during and after World War II laid the groundwork for digitized information and digital communication. Shannon's model broke information down to its simplest possible form, the bit, which represents a simple binary logic: yes/no, on/off, 1/0. A series of bits could theoretically represent any piece of information, but by using a binary system, information could be materialized, manipulated, and transmitted using electricity (switches, transistors, gates) and later magnetism (tape media and hard drives).

[24] Except for the occasional net neutrality efforts that make headlines (see Wortham, 2011, etc.).

Such a system also divorces meaning and context from information. Bell Laboratories completed a significant amount of consulting work for the U. S. government in the 1940s, and many of its developments were spurred or influenced by the war and its aftermath (Gleick, 2011). Information theory's flexibility offered new ways to communicate securely, and research using the theory was fostered by the corporate-government relationship at Bell Labs, including Shannon's (1949) subsequent work on cryptography (Slater, 1987).

As the Cold War set in as the dominant world paradigm of the 1950s, the applications of information theory continued to be developed as computer technology, spurred by the invention of the transistor and subsequent integrated circuits, became faster and more advanced. Packet switching, a key concept to computer networking and the Internet, was perhaps the most important development, conceptually combining information theory and network theory. Packet switching enables a message or piece of information to be broken down into small packets of bits, which are then sent through a network separately and reconstituted at their destination. These packets take different routes through the network, each dictated by the overall resource demands on each node in the system at the time of transmission; at every step, the packet tries to get closer to its destination, but if a closer node is busy or incapacitated, the packet will find another route (Figure 3). The important implication of a packet switching system is that it creates a distributed network, less reliant on centralized nodes or control centers. Terranova (2004) argues that packet switching is a unique characteristic of the Internet, different "from other modern decentralized media (from telegraphy to radio). … Messages are not beamed or transmitted through a channel, but broken down and let loose in the network to find their destination" (p. 64).

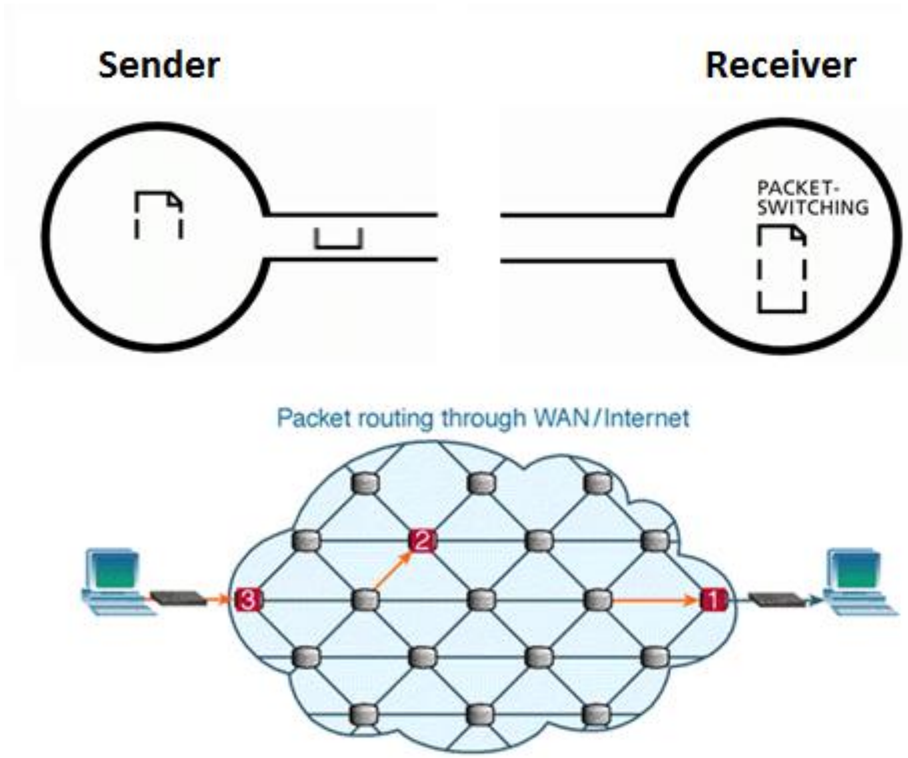

Figure 3. Packet switching (top) and packet routing (bottom). Images from Bilgil (2009) and Boerner (2010).

Packet switching was developed in the early 1960s independently by Paul Baran in the U. S. and Donald Davies in England, though each with different motivations. Baran, working at the government-funded RAND Corporation, was concerned by the damage a possible nuclear attack could cause to national communication systems, noting that a targeted attack could destroy a military or government communications hub and paralyze a system (Abbate, 1999). Through a distributed communication system utilizing packet switching, *survivable communication* could withstand such an attack, as information would route itself around the disabled location. Baran's system required that each node be "intelligent" enough to reroute a message, and thus each node in a communication system should be a computer, not just a telephone switch. Additionally, the transmitting and receiving computers would need to know how to break apart and reconstruct messages, and each packet would need address and control data appended to it. All of these conditions would be costly for a system (both hardware costs and resource costs), but would provide structural redundancy, the bottom-line for survivable communication. Abbate (1999) writes that “Baran’s system departed from traditional telephone company practice in ways that show the effect of Cold War military considerations on his design assumptions” (p. 17).

Davies’ work with packet switching developed under a much different impetus than Baran’s work. By the 1960s, England was facing a growing technology gap with the U.S. in the computing sector, a problem Labour politicians claimed was dragging down the economy. Davies and his colleagues attempted to boost the technology sector by focusing their research on interactive computing and the user experience. The computing power of the time made many tasks slow and arduous, but the development of time-sharing, where numerous users share computer processing time, was increasing efficiency and productivity. Davies envisioned similar benefits from a network that used packet switching to share the load of communication transmission, with the eventual commercial payout coming when computers eventually spread to more workplaces and homes. Ultimately:

> Davies’s concern with economics and user friendliness underscores the national context in which he conceived the idea of a packet switching network. Davies did not envision a world in which his proposed network would be the only surviving communications system. Rather, he saw a world in which packet switching networks would compete with other communication systems to attract and serve the business user and in which the United Kingdom would need to compete with

> the United States and other countries to offer innovative computer products. (Abbate, 1999, p. 28-29)

Digitized information and packet switching set the stage for the expansion of networks, from early local area networks (LAN), to campus area networks (CAN), metropolitan area networks (MAN), and wide area networks (WAN) that can crisscross a nation. Fundamentally, each of these represents a unique, unified network functioning with the same internal protocols. As the U.S. government-funded Defense Advanced Research Projects Agency (DARPA) began exploring the boundaries of its own network (ARPANET), originally located on a handful of university campuses, the need for inter-network communication became apparent. A number of protocols were developed and tested to facilitate communication between networks while maintaining the individuality and customizability of independent networks. DARPA employees Vint Cerf and Robert Kahn developed the Transmission Control Protocol and Internet Protocol (TCP/IP) suite that was ultimately supported by the early Internet community as the simple, flexible solution to cross-network communication (Abbate, 1999).

The ARPANET's expansion was initially slow, limited to academic institutions and government offices. By the end of 1970, thirteen sites were networked; by 1975, fifty-seven sites; and by the end of the decade, over 200 sites (Hafner & Lyon, 1996). By 1982, the Internet Protocol Suite (TCP/IP) was standardized and replaced the ARPANET's own protocols, signaling the true emergence of a broader network, the Internet. DARPA soon after decided to develop only the military applications of networking and transferred control of most of the ARPANET's current hardware and operations to the National Science Foundation (NSF). Commercial Internet Service Providers (ISPs) appeared in the 1980s and early 1990s, offering subscription access to the Internet. The ARPANET was officially decommissioned in 1990, and by 1995, the NSF's control and funding of any hardware and infrastructure was ended; the commercial Internet, as we it today, was fully self-sustaining.

Evolution of the World Wide Web

The privatization of the Internet and the diffusion of personal computers into homes and offices in the early 1990s enabled vast amounts of people to communicate online, but despite advancements in computer graphics, much of this communication was still text-based. Proprietary networks like America Online and CompuServ offered visually attractive services and user-friendly interfaces that exploited the power of personal computers, but the time was right for a unifying application that would enable information sharing across the Internet in a media-rich environment, and in effect, turn the Internet into a form of mass media. The World Wide Web, developed in 1990 by Tim Berners-Lee, Robert Cailliau and others at the European Organization for Nuclear Research (CERN), emerged as that application.

Led by Berners-Lee, the CERN team wanted to create a networked system that allowed scientists to create and share multimedia data while maintaining established Internet protocols (Abbate, 1999). Berners-Lee also envisioned a system that would realize the potentials of hypertext, a system of organization previously proposed by Ted Nelson that would link pieces of information together in a non-linear fashion, as well as the structure of Vannevar Bush's memex (Berners-Lee, 1999; Landow, 2006). To achieve these goals, the team needed to develop a group of networking conventions and technologies that would run on top of the existing TCP/IP protocols; HTTP, HTML, and URL all came out of this initiative. Hypertext transfer protocol (HTTP) was designed to facilitate the transfer of webpages between Web browsers and Web servers, with webpages written in hypertext markup language (HTML), a flexible system of design that could accommodate many different data formats. To find pages on the Web, a standard addressing format, the uniform resource locator (URL), was created that allowed HTTP to communicate with other Internet services like the file transfer protocol (FTP) and Usenet. The entire development process was guided by values learned from the Internet's maturity to that point: decentralization and flexibility, open architecture, and active user participation (Abbate, 1999; Hafner & Lyon, 1996).

CERN released its Web software in 1991, and by 1993, Web browsers like Mosaic were popularly used for communication across the Internet, prompting the rapid development of commercial browsers. The technical advancements of the Web opened up the social possibilities of the Internet in numerous ways. Abbate (1999) argues:

> The Web would fundamentally change the Internet, not by expanding its infrastructure or underlying protocols, but by providing an application that would lure millions of users. … Instead of being seen [as] a research tool or even a conduit for messages between people, the [Internet] took on new roles as an entertainment medium, a shop window, and a vehicle for presenting one's persona to the world. (p. 213-214)

Berners-Lee (1998) himself hoped the Web could become "a realistic mirror (or in fact the primary embodiment) of the ways in which we work and play and socialize … [and help us] make sense of what we are doing, where we individually fit in, and how we can better work together." He founded the World Wide Web Consortium (W3C) in 1994 to maintain technical standards that would promote the Web's growth.

Digital scholars have begun to map out the evolution of the Web, breaking its history into versions that each coalesce around a guiding principle (Figure 4). The early years of the Web were characterized by the *connectivity* enabled by the original development of Web technology and the advances in browser and server software through the 1990s. This period witnessed a dramatic increase in the number of people connected to the Internet via personal computers, weaving through the Web of hyperlinks and establishing a base of computer literacy.

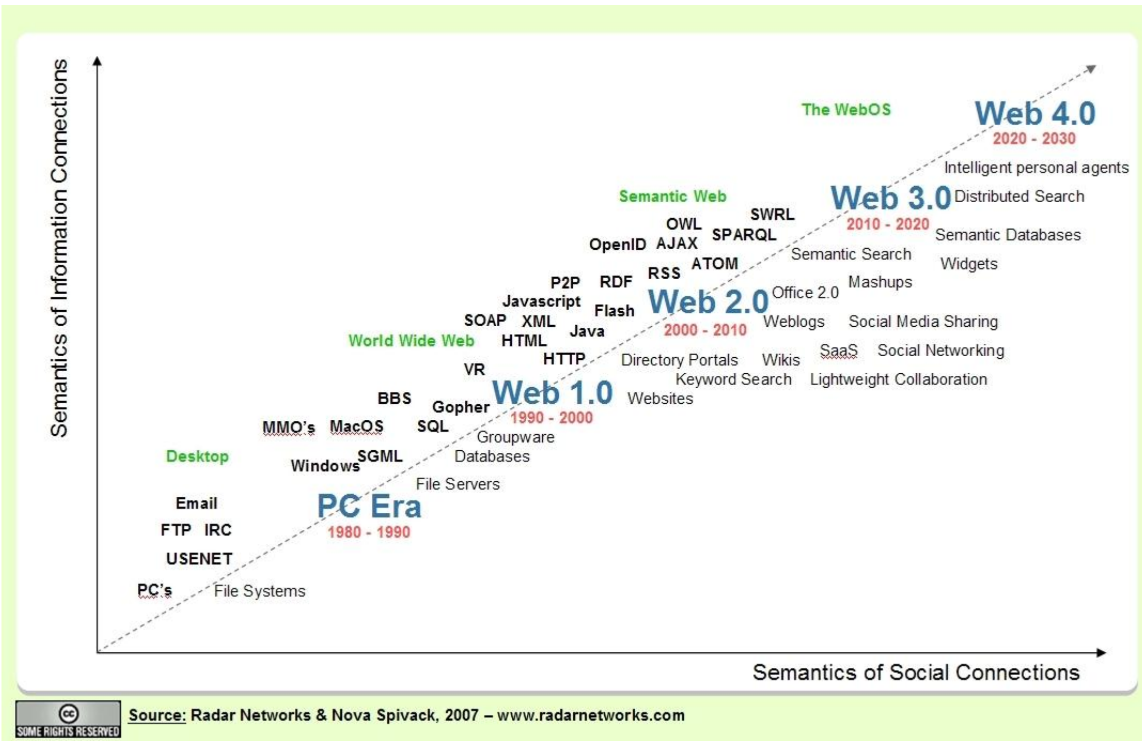

Figure 4. Evolution of the World Wide Web. Image from Radar Networks and Spivack (2008).

The second Web era, present today and popularly dubbed “Web 2.0” (DiNucci, 1999), has been characterized by *interactivity*. Cheaper and more powerful computers, faster connections, commercialization, and a new, “native” generation of users have facilitated the development of greater information sharing and collaboration abilities on the Web. In addition, the simple definition of “user” has morphed into “produser” (Bruns, 2008) or “prosumer” (Tapscott & Williams, 2006), representing the highly participatory nature of the Web where users both produce and consume content (Jenkins, 2006). Social networking, blogs, and wikis are emblematic of this paradigm. Of course, the spirit of interactivity in the Web 2.0 era was a major element in the original design of the Web, and Berners-Lee, among others, has criticized the commercial propagation of Web 2.0 jargon (Laningham, 2006).

The next stage of the Web is largely predicted to be a *semantic* Web based on natural language. As search engines continue to advance, both on the backend via advanced algorithms and database structures and on the frontend via user-friendly interfaces, the Web is expected to handle natural-language processing, making it more intuitive, accessible, and flexible (Por, 2008; Spivack, 2008). Indeed, semantic computing has become a major research area in artificial and machine intelligence, complimented by voice and spatial recognition technologies that may allow users to interact with the Web away from the screen and keyboard (Luger, 2005).

A Technical History of Wikipedia

As outlined above, the evolution of digital communication technologies, and specifically the Internet and the Web, have been shaped by their sociopolitical and socioeconomic environments, while in turn opening up new possibilities for social and cultural formations. The story of Wikipedia's origins and subsequent growth is no different; both wiki technology and the Wikimedia sites themselves have developed in a rich context of social, cultural, and economic forces. These forces both direct and constrain Wikipedia, but do so from within their own contexts. Indeed, as Bijker (1995a) reminds us: "A theory of technical development should combine the contingency of technical development with the fact that it is structurally constrained; in other words, it must combine the strategies of actors with the structures by which they are bound" (p. 15). For Wikipedia, these actors operate within the dueling ideologies of postindustrial capitalism and the open-source technology movement, each of which has shaped the site in their own traditions. The following history traces the actors, forces, and traditions that have influenced Wikipedia's development as a digital media platform.

A typical Wikipedia article page[25] will be used as a guide through this sociotechnical history (Figure 5). Articles exist in the main namespace of the site, where the most current version of the content is displayed. Often, this is the only part of the article that general visitors to the site use. However, each article has a series of tabbed links near the top of the page that allow users to engage with both the content (A) and other users (C).

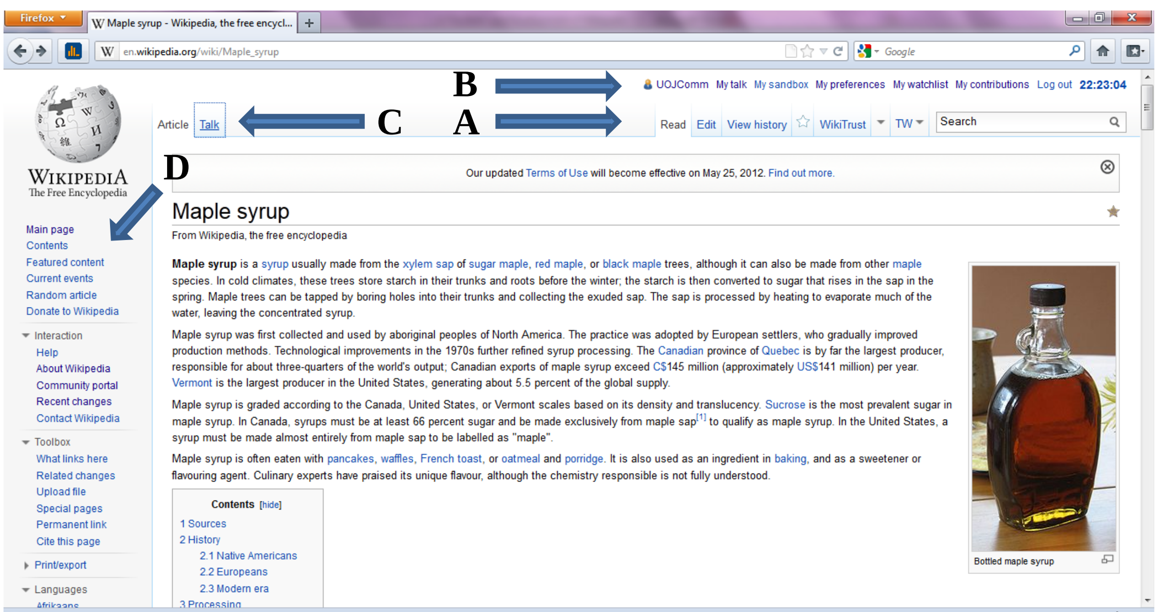

Figure 5. Wikipedia article page.

[25] Note that all of the features and functions noted here are consistent across pages in Wikipedia's main article namespace, but some variations do occur on pages in other namespaces, and special circumstances are not addressed here.

Perhaps the most fundamental of these elements is the Edit link, which brings the user to a new page featuring a scrolling textbox (E) that contains all of the article's current content (Figure 6). Here a user can add, change, delete, or revise the article's text, graphics, and multimedia features. The site uses a simplified programming language known as *wiki markup*, which is more akin to natural language than standard Web languages like HTML. Links to other Wikipedia pages are created by placing double brackets around a word or phrase (F), a process known as *free linking*. Once a change is made, the user can preview the content, submit an edit summary describing the change, mark the change as "minor," and/or submit the change, which will immediately be reflected in that article.

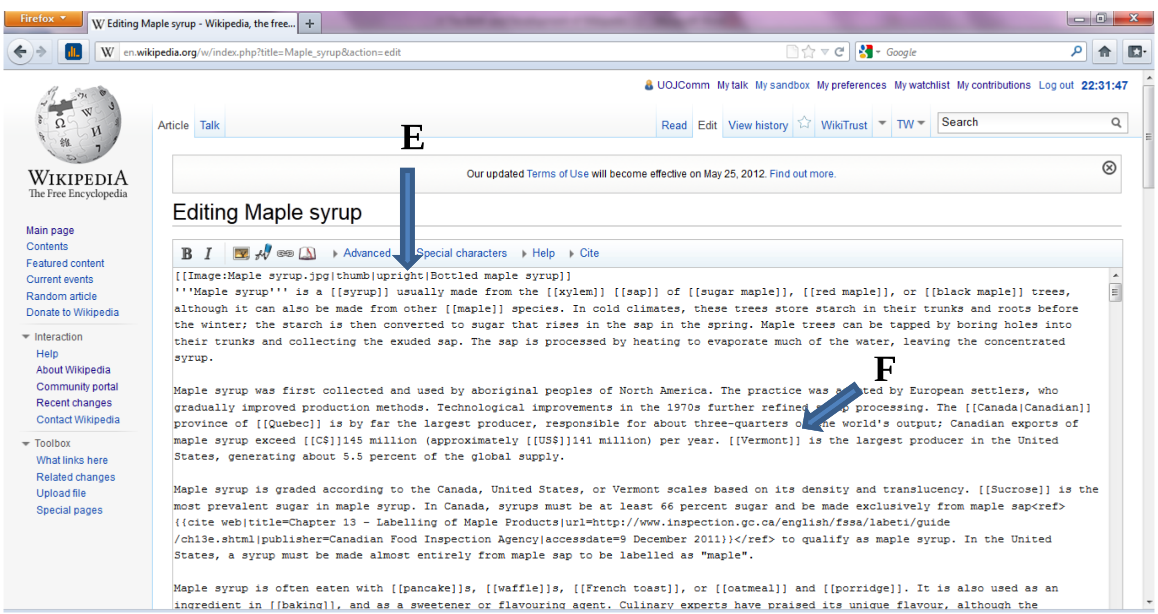

Figure 6. Wikipedia article Edit page.

Wikipedia documents and archives all changes to all content, which forces contributors to be transparent in their actions and allows for the quick and easy reversion of undesirable alterations. The View History page displays the date and time, contributor name,[26] size of the change, and edit summary for all modifications to the article (G), as well as provides the ability to undo a change and compare any two previous versions of the content (I) (Figure 7). The data on this page can be sorted in a number of ways and provides links to external tools for further analysis of the article's evolution (H).

Figure 7. Wikipedia article Revision History page.

In addition, every main namespace page has an associated discussion, or Talk, page (C) associated with it in a separate namespace (J) (Figure 8). This page provides the forum for discussing and debating any major or potentially controversial changes to an article. Talk pages also serve as a social space for editors to communicate and collaborate, though contributors are encouraged to keep the dialogue pertinent to the article; personal conversations are more appropriate for the Talk pages associated with each editor's User page.

[26] All contributors show up in the History of a page, including registered users, registered bot users, and anonymous IP users.

Logged-in, registered users also have access to a number of additional links and tools from any page, including the "★" tool, which adds that page to the user's watchlist (A). Links to the user's personal User page, personal Talk page, previous contributions,

Figure 8. Wikipedia article Talk page.

and preferences are maintained at the top of the page (B), while a number of others, including a link to the Recent Changes feed and links to all other language versions of the article, are displayed along the left sidebar (D). Finally, near the bottom of each article are listed the user-defined categories for that content (K), as well as the date of its last modification (L), its Creative Commons licensing information (M), and its Wikimedia and MediaWiki icons (N) (Figure 9).

While the aforementioned features of Wikipedia's functionality have been fairly stable since 2003, with only minor changes from MediaWiki upgrades evident to the general user, the early years of the site, the software, and the organization's technical infrastructure were quite volatile. The following history, which stretches back well before the idea for Wikipedia was even hatched, both elucidates the development of the online encyclopedia as we know it now, and establishes the historical precedent for understanding Wikipedia as a sociotechnical system.

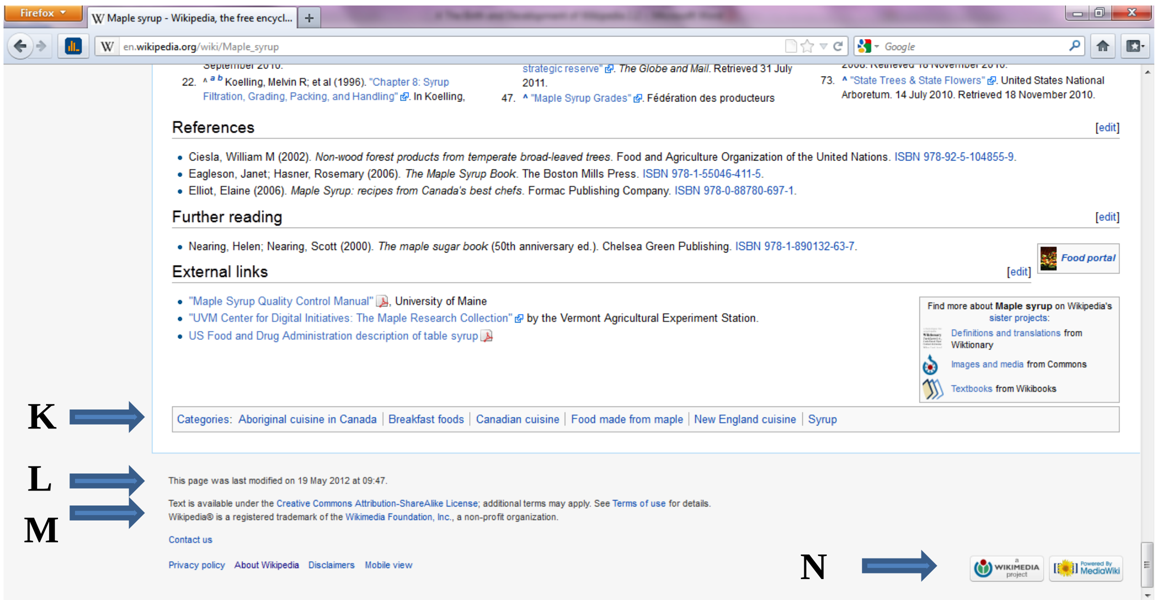

Figure 9. Wikipedia article page (bottom).

Bomis and Nupedia

Wikipedia cofounder Jimmy Wales, a former stock trader and Ph.D. student in finance, was an early Internet user, witnessing both the emergence of the commercial Web and the growth and success of the free and open-source movement. In 1996, he founded Bomis with partner Tim Shell, a search engine that found its niche by providing "guy-oriented" and adult listings and content. The site generated revenue primarily through the sale of advertising and a subscription-based photo content service. Initially Bomis constructed its own Web database, but by 1998, it was taking advantage of open content pursuits like the Open Directory Project, a volunteer-created Web listing with a copyleft license (Lih, 2009). Bomis made two technical contributions to the growing Web community at the time: the Bomis Browser and "rings" of content. The Bomis Browser was an application that allowed users to share their browsing history and patterns with the online community to improve the search engine (Martin, 1999). This data was then used by Bomis to create "rings" of associated content that any interested user could follow. Each of these functions—user feedback and categorization—would become hallmarks of the interactive Web era and essential elements to Wikipedia.

The financial success of Bomis allowed Wales to begin a pet project, an online encyclopedia. The idea was to create an expert-written encyclopedia to be distributed for free on the Web and eventually funded by advertising revenue. In January of 2000, Wales hired acquaintance and philosophy Ph.D. student Larry Sanger to oversee the project and act as editor-in-chief. The site, named Nupedia in homage to the open-source GNU operating system, went live on March 9th of that year, with its first "finished" article not published until September. In an interview about the project, Sanger emphasized that Nupedia content would be both carefully and impartially crafted by a body of volunteer academics and experts (Gouthro, 2000). The editorial process Sanger implemented, though, was painstakingly slow and alienated many contributors. The dearth of content on the site brought little ad revenue, and the project quickly stagnated; as Lih (2009) describes, "Nupedia was too much process, too little volunteer output, and not enough money" (p. 41).

Technically, Nupedia collaboration initially relied on mailing list exchanges, an inefficient way to pass content through a lengthy editorial process. By the fall of 2000, a programmer was hired to develop a Web interface for the volunteer authors to interact online more directly, but "it proved to be no better than the old method" (Lih, 2009, p.

41). Without a working technical infrastructure, and with Bomis fighting the dot-com market downturn, Wales reconsidered the future of the project.

The Wiki Way

The lifeline (of sorts) for Nupedia was an increasingly popular Web technology known as the wiki. A central new media object of the interactive Web, a wiki is a website that allows users to collaborate on content and document changes to that content using a Web browser (Ebersbach, 2006). At the same time, a wiki is a hypertext system that builds a database of pages through linking. Original wiki developer Ward Cunningham (2002) describes a wiki as "the simplest online database that could possibly work," highlighting the minimal text syntax used to create and edit content on wiki pages.

The story behind the first wiki is itself a lesson in sociotechnical development. Cunningham, a software developers for the high-tech firm Tektronix, sought a better documentation tool for collaborating programmers, "a way to document the people, ideas, and projects within the company, so people across the organization could share in that knowledge" (Lih, 2009, p. 46). Of the need he wrote: "I was frustrated that computer hardware was being improved faster than computer software. I wanted to invent some software that was completely different, that would grow and change as it was used. That's how wiki came about" (Redden, 2008). Certain functionalities of a new collaboration tool would be necessary, including ease of access by a multitude of users, automatic archiving to document past changes and revisions, the use of hyperlinks to avoid redundant material, and a simple page design to aid navigation and focus attention on content. While working at Tektronix, Cunningham experimented with HyperCard, an Apple application that could build a database of content in the form of hyperlinked digital notecards (Lih, 2009; Reagle, 2010). He decided to aggregate the functionality of different types of cards into one page, and he expanded the HyperCard convention of only hyperlinking to existing content by adding links to new, undefined pages that the user could then create. The result was a database that said to the user "Okay, I can't tell you … you tell me," which Cunningham describes "makes for top-down editing which is a good fit for describing something where you're not sure where the boundary is" (Computer History Museum, 2006).

Cunningham put his ideas into action with WikiWikiWeb, a website launched in 1995 for the Portland Pattern Repository to aid programmers in sharing ideas (Reagle, 2010). WikiWikiWeb was also the name of Cunningham's original software to run the site, written in the Perl programming language (Leuf & Cunningham, 2001). The term *wiki* comes from the Hawaiian expression "wiki wiki," meaning "very quick," and was inspired by Cunningham's memory of the Wiki Wiki Shuttle at the Honolulu International Airport (Cunningham, 2003). In describing his choice of "wiki wiki" to an editor of the *Oxford English Dictionary*, he explains the symmetry between the name and his programming syntax:

> I chose the word wiki knowing that it meant quick. I also knew that in Hawaiian words were doubled for emphasis. That is, I knew that wiki wiki meant very quick. I thought this doubling was appropriate for my technology's name because I used unusual doublings in my application as formatting clues: double carriage-return = new paragraph; double single-quote = italic; double capitalized-word = hyperlink. My program was also very quick. (Cunningham, 2003)

Ultimately, the term has become synonymous with the ease and speed in which a user can edit a wiki page.

More than just a technology, though, Cunningham's concept and implementation embodied ideals from the free software movement, where freedom implies "free of cost, free of restrictions to change and modify any content, free to redistribute, free for anyone to participate, and free of commercial influences" (Ayers & Yates, 2008). This movement was in many ways founded with the development and free distribution of the UNIX operating system by a group of Bell Labs employees in 1969, but it was formally declared with the launch of Richard Stallman's GNU Project in 1983 (Campbell-Kelly, 2003; Kelty, 2008). Representing more than its name, however, free software "exemplifies a considerable reorientation of knowledge and power in contemporary society—a reorientation of power with respect to the creation, dissemination, and authorization of knowledge in the era of the Internet" (Kelty, 2008, p. 2). Cunningham (2002) believed the simple concept of the wiki could have "profound" effects on the development of the Web, then still in its connectivity phase, and could continue in the free software tradition; of this he wrote, "Allowing everyday users to create and edit any page in a Web site is exciting in that it encourages democratic use of the Web and promotes content composition by nontechnical users." Lessig (2004; 2006), Zittrain (2008), Benkler (2006) and others would later herald such equality and interactivity as essential to the ongoing development of the Internet as a public good.

Wikipedia would become the most recognized application of Cunningham's technology, but wikis are now used in a number of settings to manage content and enable collaboration. Wikis are used both for open, Web-based content management systems and for closed group work (Ebersbach, 2008). Government agencies like the Central

Intelligence Agency[27] and the United States Patent and Trademark Office use wikis to collect, share, and analyze information, while private and corporate groups use wikis for internal documentation and project management, among other tasks (Majchrzak, Wagner, & Yates, 2006; Sipress, 2007). Wikis are also finding applications in the fields of education, medicine, and law (Leuf & Cunningham, 2001).

The Birth of Wikipedia

Wales and Sanger each claim to have come across the wiki concept and Cunningham's WikiWikiWeb independently, Wales being referred to the site by a Bomis employee, and Sanger by an old friend (Lih, 2009). Sanger's later departure from Nupedia and Wikipedia in 2002 created a rift between the former colleagues that still lingers today, so the full story of their turn to the wiki remains contested. The contributions that the wiki would make to their pursuit of a universal online encyclopedia, however, are clear; the wiki would change the direction of Nupedia and ultimately lead to its demise, as its sister site, Wikipedia, exploded.

On January 10, 2001, Wales installed UseModWiki, "an implementation of the original Wiki concept created by WardCunningham," on a Bomis server (Adams, 2007; Lih, 2009). He and Sanger considered this an experiment to hopefully improve the collaboration and editorial processes of Nupedia, and Sanger (2001a) sent a message out to Nupedia authors via their listserv that day, entitled "Let's make a wiki." In that introduction, he highlighted:

> No, this is not an indecent proposal. It's an idea to add a little feature to Nupedia. Jimmy Wales thinks that many people might find the idea objectionable, but I think not.

[27] The CIA operates Intellipedia, an online, wiki-based data sharing system dubbed "a classified version of Wikipedia … [that] is transforming the way U.S. spy agencies handle top-secret information" (Calabresi, 2009).

> This is the ULTIMATE "open" and simple format for developing content. We have occasionally bandied about ideas for simpler, more open projects to either replace or supplement Nupedia. It seems to me wikis can be implemented practically instantly, need very little maintenance, and in general are very low-risk. They're also a potentially great source for content. So there's little downside, as far as I can see.
>
> We would not integrate the Nupedia wiki into the rest of Nupedia. … It would be a completely separate part of the website. … On the front page of the Nupedia wiki we'd make it ABSOLUTELY clear that this is experimental, that Nupedia editors don't have control of what goes on here, and that the quality of articles, discussion, etc., should not be taken as a reflection of the quality of articles, review, etc. on the main part of the Nupedia website. (Sanger, 2001a)

This initial message elicited positive feedback from Nupedia contributors, but their feelings soon turned to discontent. Lih (2009) speculates that it may have been the nonconventional syntax of some wiki conventions, the informal and unstructured flow of communication, or even "the wiki's radical inclusiveness, allowing anyone into the inner circle of creating encyclopedia articles" that alienated Nupedia's academics and experts (p. 64). Sanger and Wales decided to create a new site, Wikipedia.com, for the wiki experiment, which they launched on January 15.

Nupedia continued to operate as it had, slowly forming content for "finished" articles with Sanger as editor-in-chief. He also oversaw the Wikipedia project, which quickly attracted a base of contributors from the existing online wiki community and early mentions on Slashdot,[28] and content grew quickly from the start. After only one month, Wikipedia passed 1,000 articles; by early September, it hit 10,000; and by the end of its first full year, 20,000 encyclopedic entries populated the site (Wikipedia, 2012d).

[28] Slashdot is a blog-like website featuring user-submitted content on science and technology related stories. Operating with the moniker "News for nerds, stuff that matters," the site has been known to drive large amounts of traffic to smaller sites that it links to, sometimes crippling them, a phenomenon known as the "Slashdot effect" (Terdiman, 2004).

By comparison, Nupedia produced only 24 articles (with 74 in-progress) in its three and a half years of existence (Wikipedia, 2012g).

The UseModWiki software, and indeed Cunningham's original wiki concept, was not intended to produce "finished" content for public consumption, but rather document ongoing collaboration amongst working groups. Some of the technology's features proved less than ideal for the construction of Wikipedia's articles. For example, the wiki software automatically created links when capitalized words were concatenated in the wiki markup (E), a feature known as CamelCase. This syntax was not intuitive for non-programmers, however, nor was it useful in many situations, including single words ("EncycLopedia"), or even single letters, such as the very first Wikipedia article, created on January 16, 2001, for the letter "U" (the article ultimately used "UuU" to negotiate the system). In addition, the original wiki software housed all content and discussion of that content on one page (or namespace). But some early Wikipedians felt these discussions could make an encyclopedia article unreadable and suggested a separate, linked discussion page be created for each article (Shell, 2001). Within the first year, it also became evident that UseModWiki would not scale as a database of webpages at the pace Wikipedia was growing (Lih, 2009). Collectively, these issues called for software revisions and an eventual reboot.

MediaWiki

Clifford Adams, the original author of the UseModWiki software, was also an early contributor to Wikipedia, so when CamelCase became an obvious issue during the site's early days, he gladly volunteered to design a new syntax for link creation. On January 27, 2001, he proposed to the project's mailing list a new format, *free linking* (F), which used double brackets around the designated words to indicate a link (Adams' original examples include [[George W. Bush]], [[China-Soviet Relations]], [[Physics]], [[music]], and [[Year 2000 bug]]) (Adams, 2001). The new syntax eliminated CamelCase and allowed spaces, numbers, and punctuation to be part of links, making articles flow like ordinary prose. The new code was enabled on the Wikipedia servers on February 19, and within a year support for CamelCase was dropped from the software entirely and existing CamelCase terms were converted to free links (though many usernames still carry the CamelCase legacy) (Wikipedia, 2011b). "CliffordAdams is a hero to me for helping us with this project, and for his excellent software" wrote Jimmy Wales in retrospect (Wikipedia, 2009a).

Adams's fix proved to be an elegant solution to a potentially sizeable usability problem for Wikipedia, and free linking is still used on the site today. His software, however, proved not to be the most elegant solution for a project of this nature; the site's rapid growth during its first year, both in content and contributors, exposed other facets of UseModWiki's design that did not seem to fit the project's development, from its single-page of content and discussion on the front end, to its lack of a real database structure on the back end. The Wikipedia community became worried that increased traffic driven by a growing amount of media coverage would cripple the site (MediaWiki, 2012d).

During the summer of 2001, German Wikipedian Magnus Manske began work on new software specifically for Wikipedia, based on the source code from UseModWiki. Known as Phase II (with UseModWiki retroactively dubbed Phase I), Manske's software was written in PHP to improve browser performance and used a MySQL database to store and retrieve data. Both PHP and MySQL are free and open-source, thus keeping "the spirit of [open-source] present both in the software and the content" of Wikipedia (Lih, 2009, p. 74). Phase II was adopted by the English WP in January of 2002, followed by other language versions over the next two years (MediaWiki, 2012d).

Phase II introduced many new features that specifically addressed usability issues and empowered Wikipedians to monitor the project's growth. To address the issue of crowded and confusing articles that included both content and running discussions about that content, namespaces were created. The main namespace would host an article itself, while an associated Talk namespace (C) (J) would act as a discussion board about that article. Every article would have an associated Talk page, whether it was used for ongoing, active debates or never even touched. Lih (2009) writes, "The Talk page was an innovation that was quite different from the original wiki concept, but it drew praise from Ward Cunningham. He was fascinated to see his creation adapted for use by Wikipedians" (p. 75). Other namespaces were created as well, including the User namespace for registered contributors, and the Wikipedia namespace for site policies and guidelines, each with their own associated Talk namespaces. Namespaces can be identified by the prefix amended to the page's title; for example, the Talk namespace page for the article "Maple syrup" is "Talk:Maple syrup" (J). The division of content across namespaces has become a defining characteristic of the Wikipedia wiki, which currently supports 22 different namespaces (Wikipedia, 2012aa).

Other new features introduced in Phase II were user watchlists (D), allowing each user to monitor the changes to any article or namespace desired, and display options such as skins (B), allowing each user a choice of Web designs for interacting with the site.

Despite Manske's improvements to the wiki software, though, Wikipedia continued to suffer performance lags, so other developers began working with Phase II's open codebase. By the summer of 2002, programmer Lee Daniel Crocker had rewritten the Phase II code: "I redesigned the database schema and wrote a new PHP codebase from scratch to be more efficient, though I copied the visual design and many ideas from Magnus's code" (Wikipedia, 2011a). Crocker's code was installed on Wikipedia's single server in July and became known as Phase III. The main features and layout of Phase II were maintained, but Phase III introduced additional features for both content and control. A file upload system was introduced to add images and sounds to articles,[29] and a user email system was enabled, but perhaps most significant to the development of the site was the implementation of *diff* comparisons (I) (Wikipedia, 2011a). An established convention used by programmers to compare code, a Wikipedia diff compares two specific revisions of an article to highlight changes, giving editors an important tool for both recognizing improvements and spotting vandalism. Phase III was further refined and expanded by other volunteer developers and began to be used on the Web by other projects. In July of 2003, it was officially renamed MediaWiki (N) to distinguish it from the WMF, as the software was "used in at least several non-encyclopedia contexts and will likely be used by many more in the next several years" (Mayer, 2003).

MediaWiki's core software has remained relatively stable since 2003. Despite the persistence of outages and downtime, developers at that time decided the software's architecture was good enough to move forward with "iterative improvements" rather than a total redesign (MediaWiki, 2012d); a Phase IV was not necessary. Brion Vibber, who later became the WMF's first paid employee in 2005, became lead developer of

[29] Crocker started with song clips for The Beatles and Simon and Garfunkel.

MediaWiki and coordinated subsequent releases of the software. The MediaWiki code itself is constantly being modified to correct bugs and updated to improve usability and incorporate new features. New code is integrated on the Wikimedia sites on a variable schedule, whenever a stable improvement is ready. At the same time, MediaWiki, a free and open-source program, is used on other sites across the Web, as well as by private organizations. Full MediaWiki version upgrades for third-party users roughly followed a quarterly release schedule "to get those features into the hands of other people a lot sooner" than a traditional yearly release (Vibber, 2006) (Figure 10).

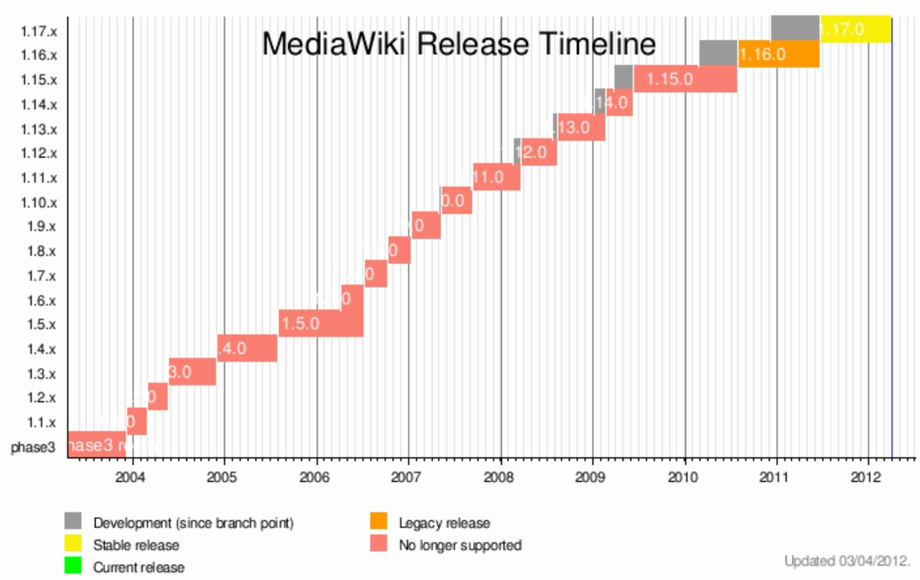

Figure 10. MediaWiki release timeline. Image from Template:Timeline MediaWiki

Some major features and functionalities have been integrated into MediaWiki upgrades, yet the software has remained flexible by being quite selective of what becomes part of the central code. WMF Technical Communications Manager Guillaume Paumier described the development process as carefully overseen by the organization:

> As with many free software communities, the community of developers discuss the features that go in or out. Large changes are discussed on the mailing list, and

> smaller changes are usually handled directly in code review comments. This is for general MediaWiki development. WMF-sponsored features are decided by the WMF's Product team, who defines priorities based on the movement's goals. [For example] this year, we're focusing on Editor engagement, and Mobile. The WMF also has a steward role in general MediaWiki development, as WMF engineers are often those who review the code submitted by volunteers. Most core MediaWiki developers are on staff, and they usually handle the release process.

Individual article editing features that have been adopted include automatically generated tables of content, individual section editing (which allows the editor to see only the wiki markup for a particular section), and system-wide features like category tags (K) and templates that make it easier for editors to organize and standardize content on the sites. Many proposed features never make it into the central code but are run as extensions, which allow additional functionality to specific MediaWiki installations without congesting the main source code that runs the Wikimedia sites.

Bots

Bots (short for "software robots") have been a reality in the world of computer science since the early 1960s, but the concept actually stretches much farther back. The first bots, developed at the Massachusetts Institute of Technology, were called "daemons," not in reference to evil spirits, but rather to Socrates' conceptualization of a non-human, intelligent, and autonomous companion (Leonard, 1997). Indeed, those early bots in the days of mainframe computing were significant helpers for computer scientists, backing up files and performing tedious and time-consuming housekeeping tasks. The first functional bot is attributed to Fernando Corbato at MIT, whose program defined some of the major characteristics of a bot's nature: bots are processes that run in the background, are normally invisible, react to their environment, and most importantly, run autonomously. Leonard (1997) claims autonomy is the "crucial variable": "[A bot] is as different from a typical software program (a word processor, say) as a clock is from a hammer. Put the hammer down, and it is useless, a dead object. The clock never stops, as long as it has power" (p. 21)

Two of the most important programs in the history of bots are ELIZA and Julia, each significant for the development of autonomous software agents. ELIZA, named after a character in George Bernard Shaw's *Pygmalion*, was developed at MIT by Joseph Weizenbaum in 1966. Programmed to mimic a Rogerian therapist, ELIZA was the first software program to successfully impersonate a human being. A user would interact with ELIZA through a text interface, typing in statements that the program would then manipulate into psychoanalytic questions (for example, upon entering "I am afraid of dogs," ELIZA would respond, "Why are you afraid of dogs?"). Weizenbaum (1976) reported that graduate student researchers would engage with ELIZA for hours and develop a sense of intimacy with the program. And though ELIZA's responses were merely algorithmic, she gave a sense of intelligence that in many ways spurred the field of artificial computer intelligence (Schanze, 2010).

ELIZA was a *chatterbot*, a term coined nearly thirty years later by Michael Mauldin (1994), a graduate student at Carnegie Mellon University and programmer of Julia. Whereas ELIZA was a local chatterbox, run from a lab, Julia was run on early networked platforms like MUDs (Multi-User Dungeons), where she interacted with anyone in the online community. Julia's conversation skills were quite advanced for a computer program; she could use humor, remember information from earlier in a conversation, delay her responses to mimic thinking, and purposefully send conversations on tangents. Even more importantly, though, Julia could answer questions about the online community that would help other users. Leonard (1997) claims, “Julia represents a giant step forward for botkind. … Julia, as a useful servant, represents in embryological form the intelligent agents waiting to be born. And Julia, as an interface, signifies the importance of the anthropomorphic approach” (p. 41-42). The implications of intelligent programs like ELIZA and Julia will be explored more fully in Chapter VI, but important to note here is the strong tradition of helper programs that vastly predates Wikipedia.

Bots were used to add articles and content to Wikipedia as early as October 2001, when entries from Easton’s Bible Dictionary were imported to the encyclopedia by a script (Wikipedia, 2012ee), but it was User Ram-Man’s automated creation, the rambot,[30] that brought bots into the consciousness of the WP community. Late in 2002, Ram-Man manually created articles for the over 3,000 geographical counties in the United States, but he decided to use his programming skills when he moved to the city and town level. Over the course of a week in October, the rambot created over 30,000 new articles on the English WP, each including consistently formatted information on location, geography, and population demographics pulled from the 2000 U.S. Census and CIA World Factbook websites. At the time, the encyclopedia had approximately 50,000 articles, so the rambot’s work expanded the project by over 60%, flooding Recent Changes and contributor watchlists (Figure 11). And although Ram-Man’s work seized upon many of Wikipedia’s early principles—“Be bold,” “Ignore all rules,” “Assume good faith”—it was met with mixed reactions on Talk pages across the site. Users Tarquin and Juuitchan wrote respectively:

[30] This bot is properly referred to as “the rambot,” including the direct article and lowercase spelling.

Hundreds of core topics are still uncovered or amateurishly-written, and here we have a page for every one-horse town across the US. It won't project a terribly good image of wikipedia; that concerns me.

And while you're at it, why limit it to the USA? Why not do England, Canada, Australia... why limit it to English-speaking countries? Why not do the whole world?? Clearly there is something absurd about this! (Wikipedia, 2012ee)

Defenders of the rambot saw these additions to the encyclopedia as a positive step:

Just linked Bitterroot to Missoula, Montana, then added to the Missoula article the fact that it is the only place that bitterroot (the state flower) grows. Took about as long as it would take to repeat an oft-made complaint against the rambot, and much more interesting, fun, encyclopedic, and productive. These articles are the foundation for the encyclopedia of the future. Use them. Improve them. (Ortolan88)

These arguments encapsulate the ideological stances of two emerging groups on the site: inclusionists, who felt the project should take advantage of its openness and include a broad range of content, and deletionists, who held a more conservative, traditional vision for the encyclopedia.

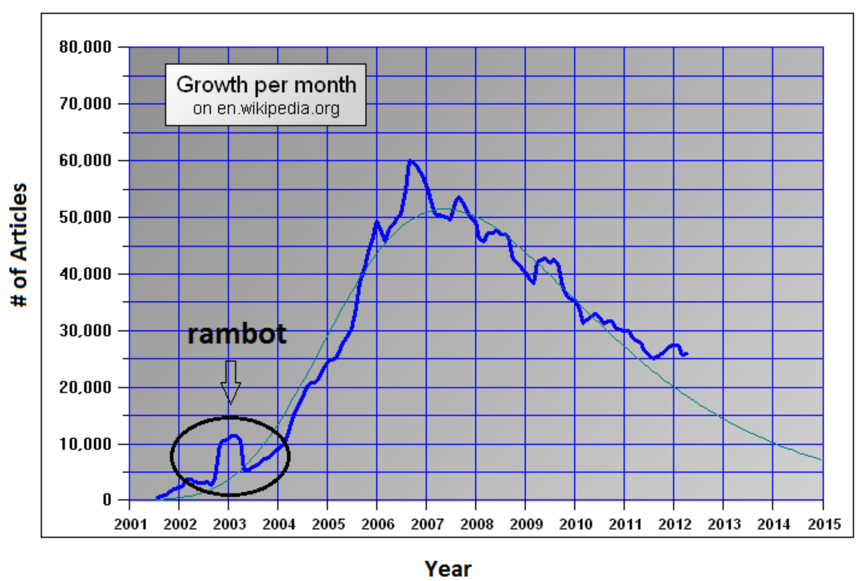

Figure 11. The rambot's effect on the growth of Wikipedia's article population. Adapted from User: HenkvD, Wikimedia Commons

From a technical perspective, critics of the rambot were worried that its speed and consistency could affect the performance of MediaWiki, which at the time was still in Phase II and run from a single server, as well as the response time of the server. At a larger level, the rambot's automated actions raised anxieties about bots on the site running amuck without the operator's awareness. Debates went back and forth regarding the implications of bot work on a theoretically "user" generated site. Ram-Man himself felt many of the arguments against bots were spurred by an "irrational fear" or technophobia among the community, concluding:

> The big issue is that people are biased against bots. No one has complained one bit about the county articles but I hear a lot of complaint about the bot added cities. I bet no one even knew that the Alabama and Alaska entries were entered by hand! The articles are almost equivalent, but people don't like one because a bot did it. (Wikipedia, 2012ee)

Ultimately, the need for a policy around bots became apparent even to Ram-Man, who drafted many of the original bot policies that remain in effect today. Bot policy on Wikipedia requires that automated edits be made from a separate account than the operator's personal account (generally with "bot" in the name) (G) and requires a bureaucrat-granted bot flag that both signifies its legitimacy and suppresses its edits from appearing on the Recent Changes page (D). Bot operators must clearly define the tasks their bots will tackle, prove the proper functionality of their bot during a trial period, and remain open and available for communication from other contributors (Wikipedia, 2012s). Early bot policy served to address the concerns of the greater Wikipedia community around automated editing, and since then it has solidified into rules and guidelines largely respected by the bot community.[31]

[31] The nuances of and controversies around bot policy will be discussed in more detail in Chapter V.

In 2006, the Bots Approval Group (BAG) was formed on the English WP to review Bot Request for Approvals (BRFA) (Figure 12). Consisting of experienced bot operators, the group would both review the soundness of a bot request and determine if there was community consensus for the task. By 2007, the BAG was facing accusations of being a technical cabal on the site, making decisions on bots without fully gauging community consensus, and adding unnecessary bureaucracy and process to the site (Wikipedia, 2012ee). But as Ram-Man pointed out: "Unlike many other Wikipedia policies that generated hot debate, the management of bots was a largely ignored and thankless job." BAG stood by the fact that the BRFA process is always open to the broader community, but few outside contributors regularly participated in the process. Opposition largely subsided, and the BAG continued their work. Today, the group consists of nine active members and a number of semi-active or inactive members.

Figure 12. Unofficial Wikipedia logos for bots and automated tools (left) and the Bot Approvals Group (right). Each demonstrates the mechanical metaphor that is often applied to bot work on the site. Images from Wikimedia Commons.

The English WP's BRFA process is the most formalized bot approval process on any version of Wikipedia. Others often have a simpler process, where a bureaucrat will grant the bot flag directly after a successful trial, barring any vocal opposition from the community. Many contributors like to work on more than one Wikipedia, however, and bot operators are no exception. Interwiki work, where content is linked across different language Wikipedias (D), has been important to the growth of many versions, and bot operators quickly found ways to automate interwiki linking. To alleviate the process of receiving permissions from every local project, Wikimedia developed the global bot flag in 2009, allowing bot access to many (but not all) Wikimedia projects. Today, the most active Wikipedia bots, including two (Luckas-bot and SieBot) with over 10 million edits, are global interwiki bots.

Hardware Infrastructure

While the MediaWiki software provides the public face of Wikipedia, a modest yet growing hardware infrastructure maintains the site on the server-side. As Ceruzzi (1998) argues, though, software and hardware are often intimately linked, jointly driving a computerized system, and the case of Wikipedia is no different. Chapter V will describe in more detail the important early coevolution of MediaWiki and Wikipedia's hardware configuration. To form a basis for that discussion, the following section describes the servers, data centers, and sociotechnical choices that have gone into the development of Wikipedia's material assets.

Servers. Both Nupedia and the original UseModWiki instance of Wikipedia were hosted on a Bomis server at its office in San Diego, California. By the summer of 2002, all Wikipedia wikis were running on a devoted server in the same location, and a year later a second server, designated as a database server, was added. These first two machines were named "pliny" and "larousse,"[32] respectively, in homage to the encyclopedic tradition that Wikipedia aimed to continue (Wikipedia, 2011a). In 2004, the

[32] Pierre Larousse (1817-1875) was "a French grammarian, lexicographer, and encyclopaedist" (Wikipedia, 2012h).

server base was expanded dramatically and the hardware infrastructure reorganized into Web servers, squid servers, and database servers. By April 2006, the WMF owned over 100 servers across three datacenters; by the end of 2007, that number rose to 350 servers; and at the end of 2011, approximately 480 servers operated its sites, with immediate plans to add 210 more machines (Vibber, 2006; Bergsma, 2007; Kattouw, 2011). For comparison, recent estimations suggest that Google owns 900,000 Web servers, Facebook at least 60,000, Amazon 40,000, and Yahoo more than 100,000 (Miller, 2011).

Even with a server population much smaller than peer top-10 websites, Wikimedia's hardware infrastructure requires consistent maintenance and remains in flux as new servers are added, old servers are repaired, and MediaWiki is updated. The WMF Operations team coordinates hardware issues, with on-site contractors at the Florida and Amsterdam data centers and a foundation employee at the Virginia data center attending to issues. Still, maintaining up-to-date, public information on servers and hardware for the Wikimedia community is difficult, according to Technical Communications Director Guillaume Paumier, as "communicating about [hardware] isn't always a priority for engineers."

Data Centers. Despite the claims that digital media and the Internet considerably reduce the importance of real-world geography (Negroponte, 1995; Rheingold, 2000; Shirky, 2008, among others), the location of computer hardware in analog space remains important for Web-based services. Though response delays are sometimes measured in milliseconds, as Wikimedia Lead Software Architect Vibber (2006) points out, "even a little bit of time can be annoying." Wikimedia developers openly acknowledge the challenges that physical barriers create for universal access, and in the spirit of their mission “to empower and engage people around the world to collect and develop educational content under a free license,” they have strategically planned to open new data centers in key global locations by 2015 (Wikimedia Foundation, 2010b). As the history of the WMF’s datacenters indicates, though, the realities of hardware infrastructure expansion are dependent on much more than a mission.

In 2004, the WMF’s hardware infrastructure was relocated to a server farm in Tampa, Florida, near Bomis’s headquarters and Wales’ home in St. Petersburg. The sites suffered a number of outages over the following year for various reasons, from blown circuit breakers and power supplies to hurricane-related power outages (Snow, 2005a). It became clear to WMF board members, MediaWiki developers, and the volunteers responsible for maintaining hardware that additional data centers were needed to decrease downtime, add data redundancy, and better serve the growing global community of Wikipedia users and contributors. Additional backup servers were temporarily hosted in Aubervilliers, France in early 2005, but the WMF’s first major non-U.S. datacenter was established in Amsterdam, the Netherlands, in June of that year. At the time, the WMF was operating on a budget under $500,000 with only one paid employee, largely lacking the resources to expand their infrastructure. The Amsterdam datacenter was made possible by an in-kind donation of 11 servers and Internet hosting by Kennisnet, a public Dutch Internet organization dedicated to primary, secondary, and vocational training (Kennisnet, 2012; Lih, 2009). The press release marking the partnership states: “People in the Dutch education system indicated that there is a huge demand for the information offered by Wikipedia and that they would benefit from improved service” (Wikimedia

Foundation, 2005). Kennisnet remains a major donor of hosting services and bandwidth for the Amsterdam datacenter (Figure 13).

Figure 13. Kennisnet Data Center in Amsterdam, 2006. The first Wikimedia rack with older servers in Wikimedia's Amsterdam cluster (left). The Kennisnet and Wikipedia logos in the Amsterdam server room (center). Fifteen new Wikimedia servers, with their front bezels attached (right). Images from Dennis van Zuijlekom, Wikimedia Commons.

Expansion continued with additional donations by Yahoo! to establish a datacenter in Seoul, South Korea (Lih, 2009). With Web servers and cache servers then in North America, Europe, and Asia, the increasing global traffic could better be handled by the Wikimedia sites (Figure 14). Additional datacenters have been added in Tampa, Florida, and the Netherlands, but the Seoul colocation went offline in 2007.

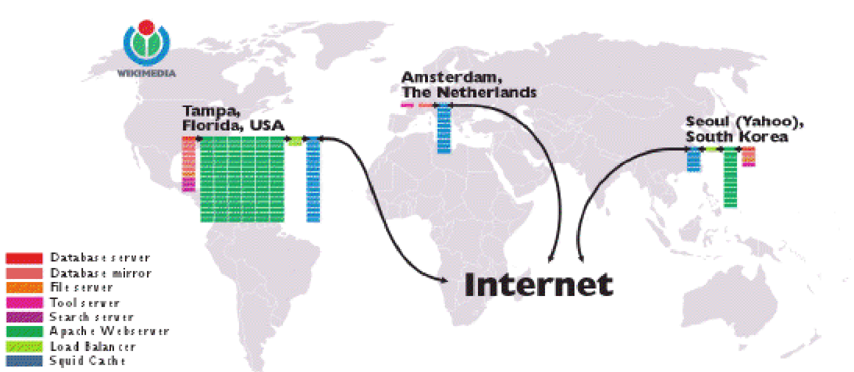

Figure 14. Wikimedia global hardware infrastructure, 2006. Image from Wikimedia Information Kit, September 2006.

Driven by the volatility of the Tampa data center, the WMF announced in 2010 plans for a new data center in Ashburn, Virginia (Wikimedia Foundation, 2010a). The MediaWiki workgroup (2011) writes: "The projects and organization are in a vulnerable position [in Florida], as any natural disaster can strike this site and cause a major impact to the availability of the projects." The new datacenter will offer proper redundancy for all hosted data, and WMF Technical Communications Director Guillaume Paumier indicated that the Ashburn facility is intended to become the foundation's primary data center in the future "because there are less hurricanes in Virginia :) [and because] connectivity is better in Virginia." Deployment of hardware at the Virginia facility began in February 2011 and is ongoing, with some Web traffic already being directed to its servers.

Moving forward, the WMF has made stabilizing infrastructure a strategic priority, arguing, "Wikipedia projects are among the most-visited sites in the world, however Wikimedia does not yet have a technological, operational and financial infrastructure commensurate with people's reliance upon it" (Wikimedia Foundation, 2011e). Their first action goal in this area is to invest in technology that will solidify current projects and support new growth, and the foundation aims to:

- Create new data centers with automatic failover features to reduce the likelihood of outages and ensure the continued operations of the site in case of catastrophic events.
- Deploy additional caching centers in key locations to serve growing audiences in Asia, Latin America, and the Middle East. (Wikimedia Foundation, 2011e)

The WMF plans to support this expansion with a revenue strategy based on small donations from individual contributors, an approach that "aligns fundraising with the rest of the Wikimedia movement" (Wikimedia Foundation, 2011e). They largely downplay funds and in-kind gifts from foundations and major donors, but the benefits from these strategic relationships seem indispensable for the level of development they envision (and will be discussed further in the final section of this chapter).

Overall, Wikipedia's early technical history was marked by a grand vision for a universal online encyclopedia and a piecemeal infrastructure of loaned equipment and volunteer labor; it's more recent history, an even grander vision of free access to free content for every global citizen and a small but organized infrastructure doing its best to keep up with growth. MediaWiki has developed from Cunningham's seminal idea about the boundlessness of hyperlinks and belief in the power of collaboration, to the open-source engine of a top-10 website that welcomes code contributions from nearly anyone. The site's technical growth and development, though, have largely been conditioned by a mesh of technical, social, cultural, and economic factors, both planned and serendipitous, as evidenced by the preceding history. To further explore some of these material and immaterial forces, the following section offers an abridged analysis of the site's political economy.

The Political Economy of Wikipedia

Mansell (2004) argues, "The relative neglect of political economy analysis in research on new media means that the overall social and economic dynamics of the production and the consumption of new media continue to be subjects of speculation" (p. 96). Some critical researchers have answered this call to action (Fuchs, 2012; Mosco, 2011; Wasko & Erickson, 2009), digging into the economic and institutional forces that have influenced the swift ascension of digital media. Often these analyses explore how new media organizations and formations fit within or diverge from established Marxist critiques of the media industries, from the Frankfurt School's (1947) notion of the cultural industries, to Murdock and Golding's (1973) ownership structures, to Smythe's (1977) audience as commodity; as such, the focus is placed on commercial media enterprises like Google and Facebook. But little research has explored the political economic forces that in many ways have influenced the development, structure, and operation of Wikipedia, the most notable nonprofit playing in a largely corporate playground.

Just as the internalist historian of technology dives into the minute details of how a tool works to uncover the forces that shaped it, the following section endeavors to look closely at the structure, funding, and labor dynamics of Wikipedia and its parent organization, the Wikimedia Foundation, to paint a fuller picture of the sociotechnical history and function of the site. Though remaining relatively small compared to its online community, the WMF has developed into a structured and strategic organization with an eye toward the future. A political economic analysis reveals the WMF shrewdly leverages its unique position in the Internet ecosystem to both benefit from relationships with substantial benefactors and maintain its aspirational mission, community of volunteer developers and content creators, and legion of individual donors.

Birth, Structure, and Role of the Wikimedia Foundation

The WMF prides itself both on being the only top-10 website to operate as a non-profit, and for keeping its formal organization relatively small. The WMF's 2010-2011 Annual Report cites 80 paid employees, compared to the tens of thousands for other top-10 website companies like Google, Microsoft, and Yahoo! And the organization highlights its size and efficiency in strategic ways, including on its donations page: "The

Wikimedia Foundation … gives everyone the tools they need to write and edit articles, ensures the servers that make the site available around the world are up and running 24/7, and runs the entire operation on a tiny budget with just a handful of staffers" (Wikimedia Foundation, 2011c).

The WMF was officially created on June 20, 2003 as a non-profit under the laws of the state of Florida, where wikipedia.com, Nupedia, Bomis, and Jimmy Wales' other projects were based at the time. In an email to the Wikipedia-1 mailing list, Wales announced that all Nupedia and Wikipedia domain names, all software and content copyrights owned by Bomis and used for the encyclopedia, and all personal copyrights used for the sites were transferred to the WMF (Wales, 2003). The two servers running Wikipedia at the time were eventually donated by Bomis as well, though not immediately, as Wales wished to obtain tax-exempt status from the IRS before transferring material assets to the new foundation. The decision to create a non-profit foundation to oversee the development of Wikipedia was an attempt by Wales to reduce his authority over the project (Poe, 2006), though he is still characterized by some in the community as a "benevolent dictator" (Reagle, 2010).

At the time of its establishment, Wales became the first chairperson of the WMF's Board of Trustees and served in that role until 2006. He was the sole board member for the remainder of 2003, but in 2004, the board expanded twice. In January, Wales appointed his business associates Tim Shell and Michael Davis to the board, followed by a summer election for two additional board members to represent users, won by European contributors Angela Beesley and Florence Nibart-Devouard (Wikimedia Meta-Wiki, 2004). The board was expanded to seven members in late 2006, nine members in early

2008, and finally ten members in 2008, when the board was restructured to "best represent the full array of community members, and … provide professional oversight for the work of the staff" (de Vreede, 2008). The new (and current) board structure includes three seats elected by the community, two seats selected by local Wikimedia chapters, one Board-appointed "community founder" seat (Jimmy Wales), and four Board-appointed "specific expertise" seats (de Vreede, 2008). Board members generally serve one or two year terms.

In November of 2007, Sue Gardner was appointed Executive Director of the WMF by the Board of Trustees after five months of serving as a special adviser to the organization. The Foundation had "determined the necessity for an Executive Director to implement strategic planning and supervise day-to-day operations of the Foundation and its Projects" (Wikimedia Foundation, 2007). Gardner previously had been the head of the Canadian Broadcasting Corporation's website and had a background in both print and electronic journalism. Her appointment signaled a shift in the organization's approach, from one of volunteer-led management to a more traditional and professional structure (Morell, 2011). Gardner's tactics were "businesslike," as she increased staff and created hierarchical structures according to their roles (Morell, 2011; UPI, 2008). Her official bio touts some of her successes at the WMF: "Gardner has more than tripled revenues, supported an 85% increase in global readership, and instituted projects and activities designed to grow readership and attract new editors" (Wikimedia Meta-Wiki, 2012h).

The WMF's mission is "to empower and engage people around the world to collect and develop educational content under a free license or in the public domain, and to disseminate it effectively and globally" (Wikimedia Foundation, 2010b). By 2009, though, the organization realized it was growing too fast to operate only on a mission and a yearly financial plan (Grigas & Roth, 2011). In 2010, the WMF published a 5-year strategic plan, the culmination of a yearlong process "to understand and address the critical challenges and opportunities facing the Wikimedia movement" (Wikimedia Foundation, 2011e). In the spirit of the project, the WMF solicited ideas and feedback from their online community, resulting in nearly 1,500 pages of content produced by over 1,000 community members in more than 50 languages. Outside voices were also solicited, including over 1,200 former community members and 65 experts and advisers, while the non-profit strategy consultancy The Bridgespan Group was hired to provide frameworks and analyze data (Wikimedia Foundation, 2011e). The fruit of this process was a strategic plan highlighting five priorities with corresponding targets for 2015: stabilize infrastructure, increase participation, improve quality, increase reach, and encourage innovation.

Assisting the WMF is an Advisory Board, created in 2006 "for the purpose of offering advice and support on a wide range of issues relevant to [the] Wikimedia Foundation … such as partnerships, public relations, financing, technology, administration, international matters and more" (Wikimedia Foundation, 2006). The Advisory Board currently has 22 members, including previous members of the Board of Trustees, prominent members of the ICT and open-source communities (including Ward Cunningham, inventor of the wiki; Craig Newmark, founder of craigslist.com; and Mitch Kapor, cofounder of the Electronic Frontier Foundation), international policy experts, and academics (including well-known Wikipedia proponent and media critic Clay Shirky).

Wikimedia Foundation Financials

Since its 2003 founding in Florida, the WMF has existed as a 501(c)(3) charitable organization in the United States. As such, the Internal Revenue Service considers the organization a nonprofit, exempt from income taxes. The WMF files a Form 990 "Return of Organization Exempt From Income Tax" annually, reporting financial data on their fiscal year (July 1 to June 30); this filing is posted on the foundation's site, along with their independently audited financial statements and a FAQ on these statements. Beginning in 2008, the WMF began publishing an Annual Report highlighting accomplishments of the previous year, changes to the Board of Trustees, future goals, and summarized financials, including a statement of activities, balance sheet, and list of major donors. Financial figures reported in this section are taken from these audited documents, unless otherwise noted.

Revenues and Expenses. The WMF has shown strong financial growth as an organization over its first eight years. Revenue has increased steadily each year, with the latest fiscal year's (2010-2011) revenue at $24.79 million, representing a 49.51% increase over the previous year (Figure 15). This figure out-performed the WMF's projections of a 28% increase in revenue (Wikimedia Foundation, 2011a). From its birth, the WMF's largest revenue stream has been individual cash donations, significantly outpacing other revenue streams, including in-kind services and donations, investment income, and earned income. In 2010-2011, individual unrestricted contributions (i.e. donations) accounted for 92.88% ($23.02 million) of overall revenue.

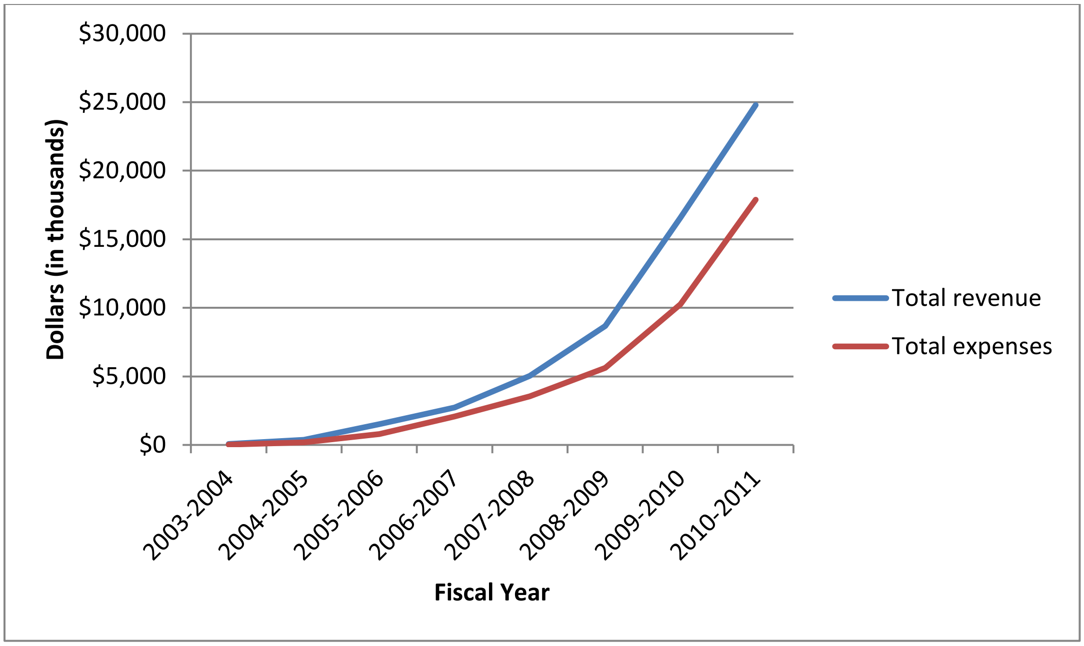

Figure 15. Wikimedia Foundation's revenues and expenses, 2003-2011. Data from [Wikimedia Foundation Financial Statements 2003-2011](#).

Expenses for the WMF have also grown steadily year-over-year, but have remained safely below revenues, allowing the organization to build a sizable cash reserve (Figure 16). Fiscal year 2010-2011 expenses totaled $17.89 million, over $2.5 million less than projected. The organization was founded with no paid employees, so the largest expenses over its initial three years of existence were Internet hosting, equipment depreciation, and in-kind service expenses. Today, the WMF has full-time administrators, technical staff, and contractors, so salaries and wages have become its largest expense, representing 40.87% ($7.31 million) of overall expenses in 2010-2011, followed by operating expenses at 32.20% ($5.76 million) and Internet hosting at 10.06% ($1.80 million).

As a new organization, the WMF saw triple-digit growth in both revenue and expenses over its first two years, but each has leveled off in the interim, with revenue growth eclipsing expense growth in 2007-2008 (Figure 16). 2010-2011 saw higher growth in expenses, mainly due to travel and expenditures associated with numerous Wikipedia 10th anniversary events around the globe.

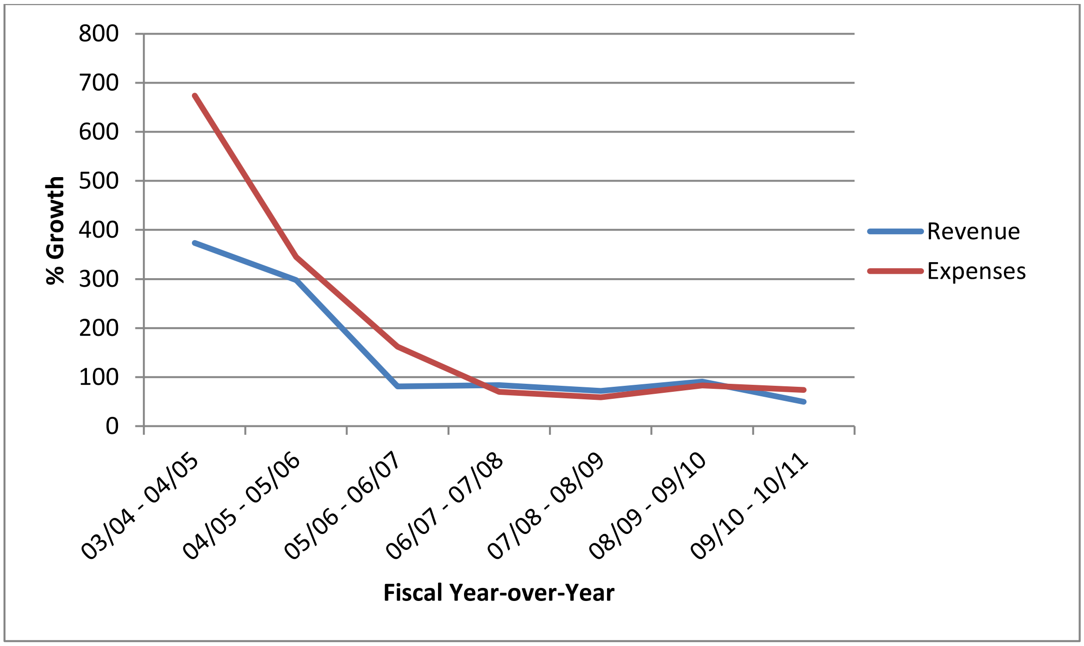

Figure 16. Wikimedia Foundation's revenue and expense growth, 2003-2011. Data from Wikimedia Foundation Financial Statements 2003-2011.

Funding. The WMF receives the vast majority of its funding from individual donors, with these contributions representing between 72.62% (2008-2009) and 92.88% of yearly revenues. The donor base has expanded dramatically in recent years, with the Foundation reporting over 1 million unique donors at the conclusion of its 2011 annual fundraising campaign, up from 573,000 donors the previous year and 130,000 donors three years prior. The organization prides itself on this grassroots funding, with WMF Executive Director Sue Gardner stating, "Our model is working fantastically well. Ordinary people use Wikipedia and they like it, so they chip in some cash so it will continue to thrive. That maintains our independence and lets us focus solely on providing a useful public service" (Wikimedia Meta-Wiki, 2011a).

The majority of individual donations come in during the WMF's annual fundraising drive, which takes place at the conclusion of each calendar year. Appeals are made through banner ads that appear at the top of content pages across WMF projects. Originally, banner ads featured simple, generic appeals featuring the WP logo or WP's cofounder Jimmy Wales. In 2010, the fundraising staff tested a bevy of alternative appeals featuring Executive Director Sue Gardner, as well as photos and messages from other WMF employees, previous donors, and average WP editors. The results indicated that the "Jimmy appeals" were most effective, but in 2011 the fundraising staff announced they were "committed to breaking the dependence on Jimmy's appeal and expanding the range of voices and faces from our community to reach our fundraising goals" (Wikimedia Meta-Wiki, 2011a) (Figure 17). The success of this strategy was measured both in the record amount of donations and the click-through analytics tracked by the WP community.

As the WMF's fundraising efforts have increased year-by-year, its fundraising expenses have followed suit. Prior to 2008, when fundraising responsibilities were solely under the purview of the Board of Trustees and the Executive Director (with the volunteer aid of the WP community), fundraising expenses accounted for approximately 6-7% of the organization's functional allocation of expenses. In 2008, the WMF hired three full-time staff to oversee fundraising: a Head of Major Gifts, a Head of Community Giving, and a Development Specialist. Other fundraising and support staff have been added since then, and as such, fundraising expenses have risen; in 2010-2011, fundraising accounted for 11.97% of the organization's functional allocation of expenses, totaling $2.14 million.

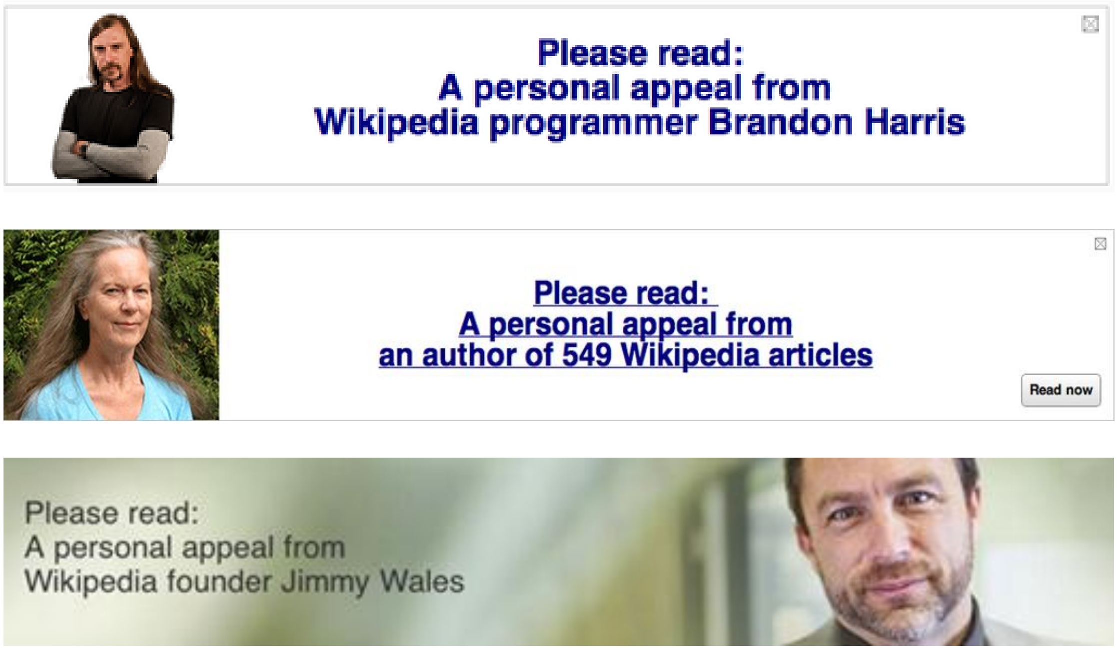

Figure 17. Wikimedia Foundation fundraising banners, 2011. Images from Wikimedia Foundation Fundraising 2011.

Though the average donation in 2010-2011 was $40.10, the WMF has a number of major donors and continuing benefactors. As early as 2006, the organization was receiving gifts in excess of $100,000 from both named and anonymous benefactors. Recurring foundational support has come from the Alfred P. Sloan Foundation, the Stanton Foundation, the Ford Foundation, and the William and Flora Hewlett Foundation, among others. Many of these major donations have been given to support specific WMF initiatives. In 2008-2009, the Stanton Foundation gave $890,000 "to reduce the barriers to participation by making Wikipedia easier to edit" (Wikimedia Foundation, 2009), and subsequent gifts from the foundation have continued to support usability initiatives, including its largest-ever grant of $3.6 million in October 2011 (Wikimedia Foundation, 2011d). The Ford Foundation has also earmarked large donations for increasing the usability of wiki technology, leading to software releases like the Upload Wizard in 2010, which offers step-by-step instructions on how to share multimedia files across Wikimedia projects.

Major corporations from the information technology sector, smaller Web-based companies, and media corporations have also supported the WMF through giving. Craigslist, Microsoft, and Google are regular contributors at the Major Benefactor ($50,000+) or Patron ($15,000+) levels, and Yahoo!, Time Warner, NBCUniversal, and Wikia (Jimmy Wales' for-profit enterprise) are among Leading Donors ($5,000+).

Local Wikimedia chapters also have a financial relationship with the WMF, including a part in fundraising efforts. Some chapters process donations from their geographical area and transfer a portion of funds collected to the Wikimedia Foundation, per established fundraising agreements (Wikimedia Foundation, 2011b). Chapters in Australia, Austria, Germany, France, Hungary, the Netherlands, Sweden, Switzerland, and the United Kingdom currently contribute to the WMF's overall revenue and are acknowledged as contributors. Some funding flows the other way as well, with the WMF issuing grants, scholarships, and event sponsorships to local Wikimedia chapters.

In addition to traditional forms of funding, the WMF receives in-kind donations of "goods and services that would normally be paid for but have been donated to use at no charge" (Wikimedia Foundation, 2011b). In-kind donations have included office space and legal or public relations services, but the bulk of these donations are in the form of Internet hosting and bandwidth. Some of these latter services are valued in the WMF's financial statements and some "cannot be reasonably estimated" and are not included in revenue figures. On average, between 75-95% of in-kind donations are valued Internet hosting and bandwidth, though the donors of these services are not disclosed. Between 2003-2008, Yahoo! and Kennisnet provided unvalued services and bandwidth, with Kennisnet, EvoSwitch, LeaseWeb , Teliasonera, AMS-IX, and Tele2 providing these services between 2009 and the present. WMF Technical Communications Manager Guillaume Paumier stated that hardware has also been contributed in-kind, but the details of those donations are not publicly documented.

The Issue of Advertising. Other than the annual fundraising drive and occasional cause-based initiative (the Stop-SOPA campaign, for example), Wikimedia sites operate free of advertising and persuasive messaging. The WMF officially maintains a stance against advertising based on ideological grounds:

> We do not believe that advertising belongs in a project devoted to education, and one that is driven by the values consistent with a balanced, neutral encyclopedia. … The current models for web advertising are also not supportive of our views on user privacy. We do not want to deliver ads to users based on their geography or on the topic they are currently reading about. … We are not against the world of online advertising, nor are we against other organizations that host ads. We just know that ads are not an appropriate thing to find in a project devoted to education and knowledge – and especially one that strives for balance and neutrality. (Wikimedia Foundation, 2012a)

And although the idea of seeing advertising on WMF sites may seem foreign today to users accustomed to the simple, ad-free design of their pages, advertising as an additional stream of revenue has been debated among board members and contributors to the sites since Wikipedia's founding in 2001. Especially in its early years, Wikipedia remaining ad-free was not a foregone conclusion.

Despite the shifting environment of Internet commerce after the dot-com bubble burst in 1999-2000, Wikipedia was born from the spirit of the emerging Internet economy. Cofounder Jimmy Wales' financial success with the search engine Bomis provided the start-up funds for Nupedia and Wikipedia, each initially run as a .com site. Fellow cofounder Larry Sanger was hired to manage and edit Nupedia, a professionally written encyclopedia that quickly became too expensive, and Wikipedia was established to utilize a user-generated content model. Sanger (2002), facing the specter of losing his paid position in early 2002, announced that Wikipedia may begin hosting advertising sold through Bomis. Many in the Wikipedia community were strongly opposed to such a plan, with the Spanish Wikipedia going so far as to create a *fork*,[33] or new website, that would host similar content without ads (Lih, 2009; Tkacz, 2011). The establishment of the WMF in 2003 as a non-profit and the switch to .org domains for WMF-funded sites solidified the project's path as a donation-based organization, and sentiment among the WMF's board and the overall Wikipedia community has been strongly against the incorporation of ads on the sites.

As both the Wikimedia sites themselves and the WMF's programs and planning have grown immensely over the past decade—perhaps more so than could be imagined during the initial years—discussions of advertising and ad revenue have inevitably remained on the radar of the Wikipedia community. Proponents of advertising largely argue that the increased income could be used to improve the technical infrastructure of the sites (i.e. faster servers, increased bandwidth, more developers to improve the MediaWiki software) and further the WMF's mission to expand free Internet access and free knowledge to underserved populations and/or underdeveloped nations around the

[33] As Tkacz (2011) describes, forking "primarily involves a split, the duplication of source code or content and the creation of a new project along with the original. The two projects proceed in different directions, but, at least initially, both draw on the original code. … As forking extended beyond its strictly computational definition to include entire projects and their contributors, it has taken on decidedly political connotations" (p. 95-96).

globe (Wikipedia, 2012o). Wales himself has acknowledged the possibilities that ad revenue could bring:

> While I continue to oppose the introduction of any advertising in Wikipedia, I also continue to agree that the discussion should evolve beyond a simple binary. I believe that if we looked at putting ads into the search results page (only), with the money earmarked for specific purposes (with strong community input into what those would be, either liberation of copyrights or support for the languages of the developing world or...). As the Foundation continues to evolve into a more professional organization capable of taking on and executing tasks (yay Sue and the growing staff!), it begins to be possible to imagine many uses of money that would benefit our core charitable goals. (Wales, 2008)

Detractors of advertising claim it would drive away core contributors who are driven by the free and open nature of the sites, as well as threaten the neutrality of encyclopedic content while congesting the wiki aesthetic that enables unfettered collaboration (Wikipedia, 2012o).

Discussions of advertising have also surfaced around speculative valuations of Wikipedia.org if it were a for-profit site, including a 2006 analysis claiming the site would be worth nearly $5 billion (Karbasfrooshan, 2006). *Silicon Valley Insider* (2009) has since backed this estimate, listing Wikipedia as the second-most valuable Internet startup behind Facebook. Calacanis (2006) calculated that advertising on Wikipedia could bring in $100 million annually even with a system where users can opt-out of seeing the ads.

For now, the WMF remains poised to continue operating without advertising revenue, but the organization does maintain relationships with online peers that could benefit a move to advertising someday. In 2011, the WMF announced an in-kind grant from Google to use their AdWords advertising service, which places sponsored links next to search results, for fundraising purposes (Signpost, 2011). Microsoft and Yahoo! remain benefactors to the Foundation as well.

Labor

Labor is a unique element to understanding the political economy of Wikipedia, as the sites currently run on the work of a small number of paid staff and a vast sea of volunteer contributors. Indeed, Morell (2011) states that this dynamic creates a "hybrid character [in] the Wikipedia ecosystem" (p. 327). Both groups have grown swiftly with the site's rising popularity, and both have developed organizational structures to manage responsibilities, provide oversight, develop gatekeepers, and negotiate control. Understanding labor relations, and in particular the division of labor between human contributors, software developers, and automated technologies, is key to understanding Wikipedia as a sociotechnical system and is considered throughout this study. This section will focus on defining and explaining some of the formal facets of labor on Wikipedia.

Wikimedia Foundation Staff. The WMF currently employs 128 workers, including 35 (27%) designated as "contractors," organized into seven departments (Wikimedia Foundation, 2012b). The Executive department is the smallest, consisting only of the Executive Director and an Executive Assistant/Board Liaison, while the Engineering department is the largest, representing over 53% of the total staff (Table 2). Engineering also houses over 77% of the organization's contractors. Overall, sixteen employees hold a title of Chief Officer or Director.

Most of the WMF's growth in staff has taken place over the last four years (2008-2012) (Figure 18). Although Larry Sanger was the first paid employee when Wikipedia was launched in 2001, his later resignation in 2002 and the establishment of the Foundation in 2003 left the organization with no paid staff. From 2003 to early 2005, the

Table 2
Wikimedia Foundation Staff

Staff and Contractors	Staff	% of Staff	Contractors	% of Contractors
Executive	2	1.56%		
Engineering	68	53.13%	27	77.14%
Community	18	14.06%	3	8.57%
Global Development	19	14.84%	3	8.57%
Finance and Administration	12	9.38%	2	5.71%
Legal and Community Advocacy	5	3.91%		
Human Resources	4	3.13%		
Total	128	100.00%	35	100.00%

Note: Figures as of 3/21/12. Data from Staff and contractors, Wikimedia Foundation.

WMF was run solely by its Board of Trustees, and the Wikimedia sites were maintained solely by volunteer labor. By January of 2005, the sites' hardware was failing to keep up with traffic growth, and the WMF purchased a "substantial" number of new servers (Snow, 2005a). Wikimedia CFO Daniel Mayer suggested a full-time employee be hired to manage and maintain the organization's hardware, a proposal met with concern from the volunteer community over "how the Foundation would handle the balance between having paid employees and relying on volunteer efforts" (Snow, 2005a). Despite the skepticism, Wikimedia community member Brion Vibber was soon hired as Chief Technical Officer, the WMF's first employee. Vibber had been a contributor to the project since January 2002, a lead MediaWiki developer, and "one of a very small number of people in the world who deeply understands the internal, technical underpinnings of [the WMF's] projects" (Danese, 2011). Later that year, Danny Wool became the second paid WMF employee, assuming the role of Executive Assistant and focusing on grant writing for the Foundation (Snow, 2005b).

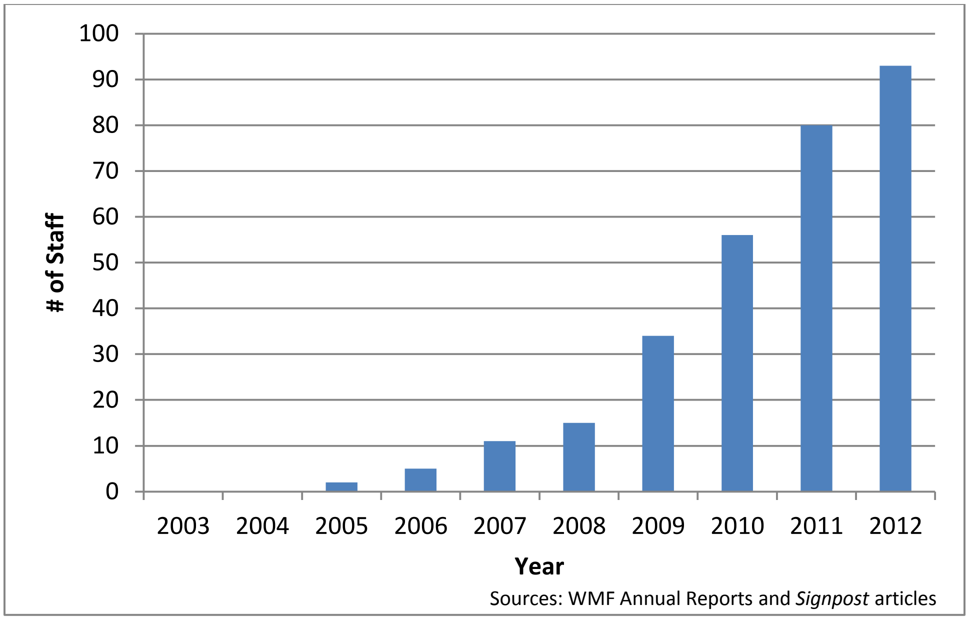

Figure 18. Wikimedia Foundation staff growth, 2003-2012.

The staff grew to five members by the end of 2006, but it was the WMF's relocation from St. Petersburg, Florida to San Francisco, California in January of 2008 that precipitated more rapid hiring. The move was a strategic one for the organization, as it hoped to take advantage of the geographical proximity to major players in the Silicon Valley high-tech sector, as well as tap into the "tech-savvy and innovative work force" (Wikimedia Foundation, 2008). As both revenues and readership continued to increase, so did hiring, and by the end of the 2007-2008 fiscal year, salaries and wages accounted for 32.43% of the WMF's overall expenses. In 2011, the Foundation reached 80 employees, accounting for 40.87% of overall expenses. According to WMF Technical Communications Director Guillaume Paumier, about half of the foundation's staff works from the San Francisco headquarters, with the other half working remotely from locations around the globe.

The WMF anticipates staff growth will continue as the organization implements its 5-year strategic plan. Projections indicate the staff will double by 2015, in line with forecasts that total expenses will more than double in that time. The majority of these new employees will focus on technology and programs, rather than governance and administration (Wikimedia Foundation, 2011e).

Volunteer Contributors. WMF Bugmeister Mark Hershberger claims that volunteers are both a strength and a weakness for Wikipedia and MediaWiki's development: "The strength is that people are working on what they love. The weakness is that you can't force volunteers to do anything." Originally an enterprise guided only by volunteer labor, today Wikipedia operates at the intersection of an organizational strategy, a small paid staff, and a legion of volunteers. Recent research has suggested that the number of active contributors[34] has plateaued and may be in decline (Ortega, 2009), and the WMF has made increasing participation around the globe a strategic priority (Wikimedia Foundation, 2011e) (Figure 19).

Volunteers are the lifeblood of Wikipedia, responsible for all encyclopedic content, as well as most of the policies, guidelines, and conventions that create the social structures on the site. Despite the relatively low rate of active editors (estimated to be .83% at the time of this writing), over 16,600,000 accounts have been registered across versions of the project, with a large, unknown number of unregistered users also contributing (Wikipedia, 2012hh). Collectively, these users make 12.4 million edits to the

[34] On Wikipedia, an "active" contributor is defined as a user who makes more than five edits per month. A "very active" contributor makes more than 100 edits per month, and data suggests that these users have plateaued as well (Zachte, 2012b).

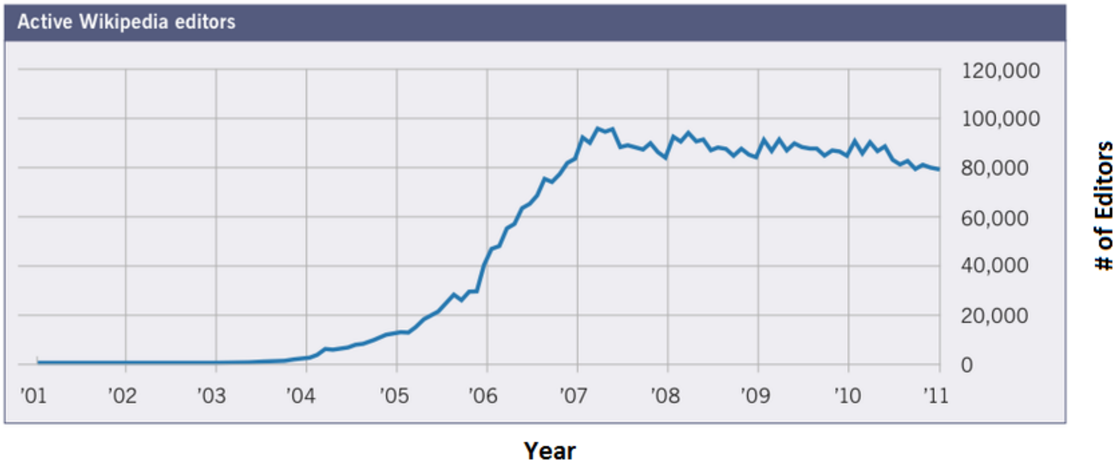

Figure 19. Active Wikipedia editors, 2001-2011. Chart from [Wikimedia Strategic Plan, 2011](#).

contributing (Wikipedia, 2012hh). Collectively, these users make 12.4 million edits to the site per month, with the content created attracting nearly 35% of the Internet population each day (Zachte, 2012a). This volunteer labor, offered largely in the spirit of the open-source movement and the desire to contribute to open knowledge, likely constitutes much of the worth represented in the site's $5 billion valuations, as it creates an enormous audience for potential advertising.

A much smaller group of volunteers contribute to the development of the MediaWiki software, extensions, and gadgets. MWF Volunteer Development Coordinator Sumana Harihareswara estimates that the technical community who interact with MediaWiki is in the thousands, with only about 75-100 of them contributing code for MediaWiki core improvements in a given month[35]. She questions, though, the definition of this collaboration: "Can people who are working on similar things but don't know each other or communicate be called a community?" Her point indicates how

[35] In January 2012, statistics on MediaWiki contributions began to be featured in the monthly Wikimedia engineering report. That month, 100 unique *committers* (individuals with the ability to modify code) contributed to MediaWiki, followed by 67, 98, and 53 over the next three months respectively (MediaWiki, 2012f).

software development for the site has largely been assumed by the WMF; most of the primary work on the MediaWiki core today is completed by WMF engineering employees and contractors, though as User MaxSem pointed out, many of the latter group begin as volunteers. Indeed, labor on Wikipedia can be understood as both fluid and stable, a notion explored more in Chapter V.

Overall, the preceding review of Wikipedia's political economy reveals a project that is in many ways far from its humble beginnings, yet insistent on managing a trajectory that does not stray too far from those ideals. As Executive Director Sue Gardner (2011) states:

> The Wikimedia Foundation is a weird organization, full of contradictions and ambiguity and messiness. We are deeply rooted in the free software / free culture movement, which is still generally perceived as 'fringe' and 'radical,' and yet we operate one of the world's most popular websites. We share attributes with large cultural and educational institutions and also with Silicon Valley start-ups. We're young, growing fast, and experimenting a lot.

In many respects, the WMF has been successful in their experimentation, establishing a niche in the highly commercialized Internet environment where they can strategically exploit relationships to maintain their non-profit ethos. In many other respects, from maximizing fundraising efforts to mobilizing their base of volunteer contributors, they are still figuring things out. Most crucially, though, they have established and are sustaining a material infrastructure strong enough to facilitate the constant and dynamic interactions between its technical and social elements.

Conclusion

This chapter has explored the various actors and forces that have influenced the technical development of Wikipedia, arguing that historical, ideological, and material elements each have conditioned the choices made to initiate and grow the project.

Wikipedia has developed many traditional organizational structures, from an administrative parent foundation with strategic goals to the managed release of software revisions; and yet, the project was born from strong ideals of openness and mass participation that are reflected in every wiki page, and these ideals continue to influence the development of its technology. Indeed, although Wikipedia is an experiment in progress, in many ways learning as it grows, it has also quickly recognized its unique position in the Internet economy, successfully managing relationships with both corporate benefactors and volunteer enthusiasts. From historical and political economic perspectives, Wikipedia has evolved a strong technical infrastructure that both enables and is a key part of its sociotechnical nature, which is detailed in the following chapter.

CHAPTER V

WIKIPEDIA AS A SOCIOTECHNICAL SYSTEM

The idea of a sociotechnical system is both plain and obvious. We know from experience that no man is an island and that life outside the laboratory is a complex milieu of forces and influences, affecting our constant interactions with the people and things in our environment. And yet, we are often driven by a thirst for simple cause and effect, a way to understand experience by isolating, then investigating. The sociotechnical systems perspective resists this impulse by reminding us of the obvious: we are not independent from our environment, and in fact we are enmeshed in a network of causes and effects that continuously condition one another.

The emergence of digital media has brought new environments and experiences to understand, and our impulse to isolate and investigate has guided much research on the Internet since its inception, including studies of Wikipedia over the past decade. Though they generally acknowledge the importance of both the *social* and the *technical* on the site, most Wikipedia research focuses either on the social world—What motivates contributors? How do they establish community? How is order maintained?—or the technical collaborative properties of the technology, without thoroughly investigating how the social and the technical constitute each other. The previous chapter demonstrated how social ideals, ideologies, and relationships have played a major role in Wikipedia's technical development and current organizational operations, which in turn raises the question: What can we learn about how the project works by exploring the details and microstructures of the site as a sociotechnical network of actors? The current chapter addresses this question, ultimately arguing that a sociotechnical perspective is in fact the best position from which to understand how the site remains concurrently stable and fluid amidst the continuous activity of millions of actors, both human and otherwise.

To do this, the chapter unfolds in three steps. First, a very functionalist account of the social and technical infrastructures of the site is presented, chronicling facts and facets of its structure and organization necessary for a more nuanced sociotechnical understanding. Building from this foundation, a sociotechnical analysis of Wikipedia is then offered to explore the ways in which human and technological actors work together and shape one another in pursuit of the project's goals. Finally, an exploratory case study using social network analysis is discussed to offer insight into how the concept of *work* is considered within such a sociotechnical system. Individually, these steps offer different points of entry from which to examine Wikipedia as a dynamic platform of digital media activity, and collectively, they advance this project's thesis that this activity is most usefully understood as a sociotechnical system.

The Social Infrastructure

The Wikipedia Community

In January 2011, the Wikimedia Foundation released a short video celebrating the tenth anniversary of Wikipedia. In the video, Wikipedia co-founder Jimmy Wales announced that "even though we've got millions of articles, even though we're in hundreds of languages, there's still a lot of work to do" (Wikimedia Commons, 2011). Wales' call to action was an invitation for those who use Wikipedia but have never contributed to become an active part of "the sum of all knowledge." The call, though, may be more than a friendly request to the larger online community. A 2009 study revealed that Wikipedia's declining editor base could be a major problem for the project going forward (Ortega, 2009), as the growth of editors and articles on the English language version may have plateaued around 2007 (Johnson, 2009).

Tracking the Wikipedia community is a difficult task for a number of reasons. Users are not required to register on the site in order to contribute content or make changes, although they are certainly encouraged to register in order to customize their editing interface, establish a reputation, and participate in the consensus process (Wikipedia, 2012gg). Even within the registered community, though, technical and procedural changes have made a global census problematic. In 2008, MediaWiki implemented a unified login system, where newly created accounts would be recognized across all Wikimedia projects. Where possible, older accounts were merged during the changeover, but over 100,000 accounts needed to be manually merged by individual users if they desired to do so (Wikimedia Meta-Wiki, 2012b). Consequently, there is now a bit of uncertainty when estimating the total number of registered Wikipedians, though a conservative baseline figure is possible.

At the time of this writing, Wikipedia (2012i) had 16,600,496 registered users, though only 138,344 (0.83%) had made a contribution within the last 30 days, one criteria for being deemed an *active user*. As many as 60% of new registered users never participate again after creating an account, though a 2009 study suggests that new users and users with low edit counts actually create more content than experienced users, who spend more time editing and formatting content (Blodget, 2009; Goldman, 2009). Demographically, surveys have shown that users are largely male (87%), young (15-49 years old), English-speaking, technically inclined, and formally educated (United Nations University, 2010; Wikipedia, 2012dd) (Figure 20).

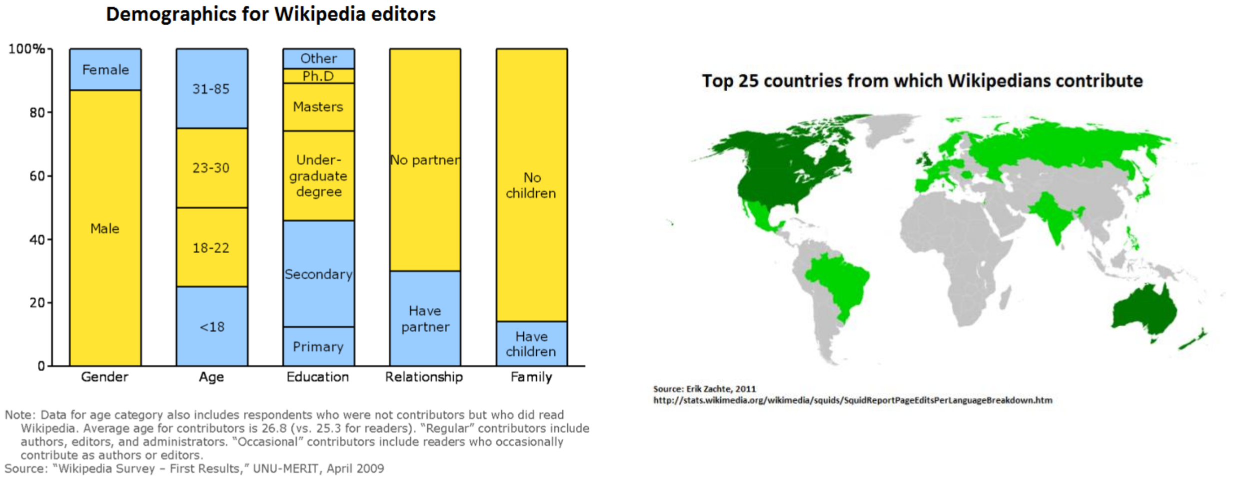

Figure 20. Demographic data on Wikipedia editors. Yellow indicates categories with a majority (left), dark green indicates nations where the majority are native English speakers, and light green indicates nations where English is largely a second language (right). Images from Wikimedia Commons.

Ortega's (2009) quantitative comparative analysis of the top-10 language versions of Wikipedia brought issues of community into the spotlight. The study found "a severe risk" of decreased contributions to these projects as "the inequality level of contributions from logged authors is becoming more and more biased towards the core of very active authors" (Ortega, 2009, p. 158). Further, trends in new contributors showed that the core of editors producing Featured Articles, the best work on the site, was not scaling with the size of the project, indicating "an untenable trend" of reliance on the biggest contributors and a prediction that these editors could burn out. The study's conclusions, along with Bongwon et al.'s (2009) study of content and contributors, brought Wikipedia's crisis of community into mass media headlines (Angwin & Fowler, 2009; Johnson, 2009), with the WMF taking notice as well. Increasing participation became a strategic priority for the organization, which wrote in their 2009-2010 Annual Report:

> Wikipedia's 'boom years' of participation are behind us. Now we face the challenge presented by a slowly decreasing number of editors. To gain new

> editors, we must adapt our editing technology to the changing web, improve the social experience for new users, and grow participation in the Global South by catalyzing on-the-ground work by the community. (p. 10)

Since then, the WMF has launched large-scale usability initiatives to study how new users interact with the wiki software, developed MediaWiki tools and features to improve the user experience, and ramped up outreach efforts to new areas of the globe, including plans for the first non-U.S. WMF office in India (Wikimedia Foundation, 2010a; Wikimedia Foundation, 2011e).

Roles and Privileges

In both appreciating wiki philosophy and recognizing that volunteers are the lifeblood of the project, one of the founding principles for Wikipedia, originally outlined by Wales, was that the social structure should remain flat and open to anyone: "There must be no cabal, there must be no elites, there must be no hierarchy or structure which gets in the way of this openness to newcomers" (Wales, 2001). As the project grew in both size and popularity, though, user roles were developed by the Wikipedia community to manage the administrative functions of the site, and role-based access control was built into the MediaWiki code to flag individual accounts with additional technical access (Wikipedia, 2012n). As Reagle (2010) comments, these "delineations of authority are suspect" for the larger community, highlighting "responsibilities rather than rights" for flagged users (p. 127). Ayers, Matthews, and Yates (2008) describe user roles as "functional distinctions, not an indication of the editor's importance on the site or the role he or she takes in contributing to content" (p. 325). Still, user roles do in effect create a hierarchy of access that have implications for the shape and activity of the community at large.

Administrators, *bureaucrats*, and *stewards* are the most important user groups to understand when considering issues of access and privilege on WP (bots as a user group will be discussed later in this chapter). Administrators, also known as *admins* or *sysops*,[36] have access to technical tools and functions of the site that are unavailable to the general community of users. Admins can protect or delete pages and block or unblock other users' accounts. Administrator rights are granted to users by the community through a formal process, including a nomination, discussion of previous history and actions, and the assessment of consensus around the candidate. As such, the decisions admins make are meant to keep the best interests of the larger WP community in mind and represent the consensus of stakeholders in any specific situation. They have a higher knowledge of policies and procedures and can answer questions, intercede in conflicts, and generally aid the work of the larger community. In addition, admins automatically receive technical privileges like "reviewer" and "rollback" which increase their ability to undo vandalism. There are currently nearly 1,500 administrators on the English WP.

Bureaucrats constitute a much smaller group on Wikipedia, with only 34 active on the English WP.[37] Bureaucrats have all administrator access rights, but in addition, they have the important power to add other bureaucrats and administrators, as well as remove admin and other access level rights and rename user accounts. Bureaucrats are also the only local user group with the ability to grant and revoke bot flags. These are "exceptionally trusted users" charged with carefully judging consensus around user nominations, but the community is careful to emphasize that they are not "super-admins" (Wikipedia, 2012ff; Wikipedia, 2012u). In fact, the term *bureaucrat* was chosen for the

[36] Short for "system operators."

[37] Only 49 users in total have held bureaucrat rights since the role was established in 2004.

role so as not to make it a status symbol: "It should be something nobody really wants—something people do because it needs doing, not because it gains them credibility and influence" (Moeller, 2007). Bureaucrats are nominated and chosen in a similar fashion to admins, and their rights are limited to the scope of the local Wikipedia on which they were granted. Bureaucrat Mbisanz feels the role is "fairly non-controversial … [and] has a substantial impact."

The final group discussed here are stewards, or users with complete access to all privileges across all Wikimedia sites. This global access is granted through a successful election on the Meta-Wiki site,[38] with a minimum threshold of 80% support by at least 30 community members required. Stewards have all the rights of administrators and bureaucrats, as well as *checkuser* and *oversight* rights, which allow them to view the IP addresses of users and hide previous revisions of content, respectively. Stewards often act as admins and/or bureaucrats on smaller wikis with few or no users in those roles, but they generally defer to local admins whenever possible. Their global rights allow them to deal with emergencies like cross-wiki vandalism, but they are expected to otherwise be conservative with their power, remaining neutral when evaluating consensus and staying away from creating policies (Wikimedia Meta-Wiki, 2012e). There are currently 43 stewards on Wikimedia projects.

Bureaucracy

Bureaucracy on Wikipedia goes well beyond delineations of user privileges and rights; policies and processes have become important to the project's functioning. With an active contributor base of more than 100,000 in a given month, the Wikipedia

[38] Meta-Wiki is an organizational site for the WMF's various projects. It's discussions "range from coordination and documentation to planning and analysis of future Wikimedia activities" (Wikimedia Meta-Wiki, 2012c).

community is too large to operate without guiding principles and agreed upon rules. The question then becomes: When does necessary order become unnecessary bureaucracy? Wikipedians vary greatly in their opinion of where this line should be drawn.

Early in the project's history, Wales (2001) declared the principles he felt were important for Wikipedia to maintain, stating, "As we move forward with software and social changes, I think it is imperative that I state clearly and forcefully my views on openness and the licenses." He argued that contributors must "Do the Right Thing" by 1) maintaining a policy of neutrality, honesty, and respect in their article writing and communication with others, 2) welcoming newcomers and remaining open both in terms of content (free distribution licenses) and community (anyone can participate), making changes to software gradual and reversible, 3) and remaining outwardly focused as a credible work of reference. Many of Wales' original points have been institutionalized by the community as the Five Pillars:

1. Wikipedia is an encyclopedia.
2. Wikipedia is written from a neutral point of view.
3. Wikipedia is free content that anyone can edit, use, modify, and distribute.
4. Editors should interact with each other in a respectful and civil manner.
5. Wikipedia does not have firm rules.

This final pillar reads:

> Rules in Wikipedia are not carved in stone, as their wording and interpretation are likely to change over time. The principles and spirit of Wikipedia's rules matter more than their literal wording, and sometimes improving Wikipedia requires making an exception to a rule. Be bold (but not reckless) in updating articles and do not worry about making mistakes. (Wikipedia, 2012b)

Indeed, as early bot operator Ram-Man pointed out, "Be bold" and "There are no rules!" defined the early spirit of the site, but today the lack of rules seems a point of semantics, as Wikipedia operates with over 200 policies and guidelines covering issues of content, conduct, deletion, enforcement, legality, and procedures. The preceding disclaimer, then, attempts to negotiate the founding spirit of the project with the order and structure that has come with its growth and popularity. As it was in the early days, nearly all collective decisions on Wikipedia still are made using a process of consensus, so theoretically anything can change; current policies and guidelines, however, also have the weight of historical precedence behind them, essentially making fundamental changes to the order of the site an uphill battle. Still, as Reagle (2010) points out, "Even in the face of a proliferation of process, the open content community values of transparency and integrity are largely preserved" (p. 91)

The confluence of structured user roles and institutionalized policies raises issues of governance that have attracted the attention of social science researchers, who largely find the decentralized authority on Wikipedia works quite well, as it reflects the ideals of the contributing community. In tracing the development of the Verifiability policy for content, Konieczny (2009) argues that the site successfully resists oligarchy, "retaining [the] idealistic goals" of its contributors. Forte, Larco, and Bruckman (2009) contend that governance on Wikipedia relies on a system of "community-generated social norms" articulated as policies, and that even as the "nodes of governance" on the site grow larger and more complex, their decentralized nature represent the embedded philosophy of Wales and early Wikipedians. Interestingly, whereas Ortega (2009) indicated the site's reliance on core contributors is likely harmful to the continued expansion of content and community, Forte, Larco, and Bruckman (2009) argue that "the continued presence of 'old-timers'" (presumably Ortega's "core") have a positive effect on Wikipedia by maintaining social norms and organizational ideals (p. 71).

As cited in the previous section, the term *bureaucrat* was selected to describe a particular user role precisely because of its negative connotation, and *bureaucracy* is often used strategically by the community to elicit a sense of over-governance. But regardless of whether it is labeled bureaucracy, governance, or process, and despite the findings of the preceding studies, many Wikipedians are wary of the culture that power and division can create on the site. Efforts that are exclusive and not transparent are dubbed *un-wiki*, as in the case of the Esperanza project, an association formed in 2005 to "indirectly support the encyclopedia by providing support and other assistance for Wikipedians in need, and by strengthening Wikipedia's sense of community" (Wikipedia, 2012w). In actuality, the project was run in private, often off-wiki in private IRC chat rooms. User Cyde characterized Esperanza as "bureaucracy-heavy. It was run like an exclusive members only club. It was very un-wiki," while User Dev920, in her nomination to delete the project, argued:

> Esperanza is thoroughly unwikipedian [sic] in its desire for endless bureaucracy. At the time of the [first nomination], Esperanza has a seven member council who held closed meetings on IRC that made binding decisions about Esperanza. Any contentious decision was to be passed up to them. There was no consensus building, no discussion, nothing. (Wikipedia, 2007a)

Ultimately, Esperanza was dissolved in 2006 with a notice posted to the project's page: "This essay serves as a notice to all editors that existing projects must be open and transparent to all editors at all times, not to be overly hierarchical lest they are to meet a fate similar to Esperanza's" (Wikipedia, 2012w).

Though but one (perhaps extreme) example, Esperanza is representative of the community's perception of bureaucracy and one way they often deal with it—head on. But many interviewees reported the damaging effect of bureaucracy in more common scenarios. Reputation matters on Wikipedia, and a good reputation can mitigate red tape in ways that reinforce inequalities. Users Anomie and Kbdank71 cite “vested” and “established” contributors who “are allowed to get away with far too much,” often receiving preferential treatment from administrators and avoiding sanctions for breaking policies. Again tied to the core/periphery view of the community, these contributors are viewed as too important to upset with sanctions, so their work is tolerated, even if it drives other contributors from the project. Unlike the Esperanza case, though, administrators will sometimes turn away from these situations in frustration: “People who are not here to make WP better should be shown the door. Unfortunately, this is not a majority opinion. So I stay away from the drama, and use my admin powers elsewhere” (Kbdank71).

Additionally, these situations, and a culture of bureaucracy in general, may be keeping talented contributors from other language versions away from the English WP. Three European Wikipedians cited bureaucracy as their reason for no longer participating on the English language version, with one describing the formal sanctioning board ArbCom[39] as “a political/bureaucratic hellhole, to be honest, [which] is in my opinion very ‘un-wiki’” (Jon Harald Søby). User Multichill worries that this emphasis on policy is creeping into cross-language Wikimedia projects like the Commons, where rules and guidelines are “scaring off bot operators who don’t feel like dealing with it.” The implications of bureaucracy, specifically around the bot community, will be explored later in this chapter, but clearly the social structures and instruments of regulation on Wikipedia play a major role in the direction of its evolution.

[39] ArbCom, short for Arbitration Committee, is a panel of Wikipedia editors that mitigates disputes between other editors on the site. ArbCom makes binding decisions for conflicts that are not resolved in normal WP community forums, and their authority is generally seen as the highest on the project.

The Software Infrastructure

Wikipedia is the largest instance of wiki technology existing today, a point of pride for both the WP community and the WMF. Overall, the project consists of more than 21 million articles, including nearly 4 million on the English language version alone. In 2010, it was estimated that a print version of the English WP would consist of 2,647 volumes, with each volume containing approximately 400 pages (Wikipedia, 2010) (Figure 21). Written by more than 16 million registered users and perhaps an even larger number of anonymous contributors, logic would suggest the need for a massive technical infrastructure to manage the size of this content and community. While the technical resources required to run the site have grown considerably since that first Bomis server,[40] the light, minimalist wiki concept has scaled remarkably well as Wikipedia has developed into a globally relevant website.

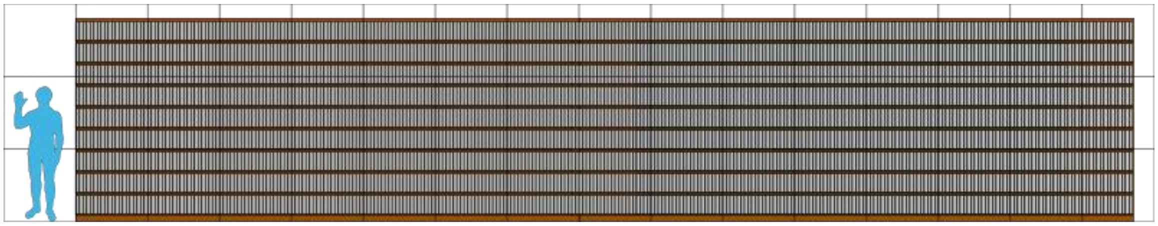

Figure 21. Estimated size of the English Wikipedia in print form, as of August 2010. Image from Nikola Smolenski, Wikimedia Commons.

So how does this work? This section will look more closely at how the wiki is implemented by Wikipedia and the Wikimedia movement into a technical infrastructure that is firm and yet flexible. The key here is to understand how Wikipedia is maintained through a layered set of technologies, from the core MediaWiki code, to customized extensions and add-ons, to software robots run from separate servers or the client-side.

[40] Wikipedia's technical history was detailed in Chapter IV.

Though a deep, technical analysis from a computer science perspective is not necessary here, a discussion of relevant technical nuances will set the stage for the chapter's subsequent argument that the social and technical infrastructures of Wikipedia are codependent and constitute a sociotechnical system.

To begin understanding these layers of technology, consider how basic interaction with Wikipedia takes place from a technical standpoint (Figure 22). The site runs from a core software package, MediaWiki, which handles most basic functions on the server-side. In order to keep MediaWiki light, flexible, and adaptable to future changes, only major features and upgrades are integrated into the core code. Smaller, more personalized, or more experimental features are made into software extensions, which work on top of the MediaWiki core and make its functionality more robust. To interact with MediaWiki, users engage with an interface. For human users, this is the graphical user interface (GUI) of the familiar Wikipedia webpage. Bots, however, interact with the application programming interface (API), or a set of specifications used by software components to communicate with each other. Both human and bot users generally interact from the client-side, an interesting similarity to note at the beginning our sociotechnical exploration of Wikipedia; bots, though essentially immaterial lines of code comparable to MediaWiki and extensions, actually occupy a position in the system more akin to the human users. Before diving into the implications of such a setup, though, let us consider other significant facets of Wikipedia's technical infrastructure.

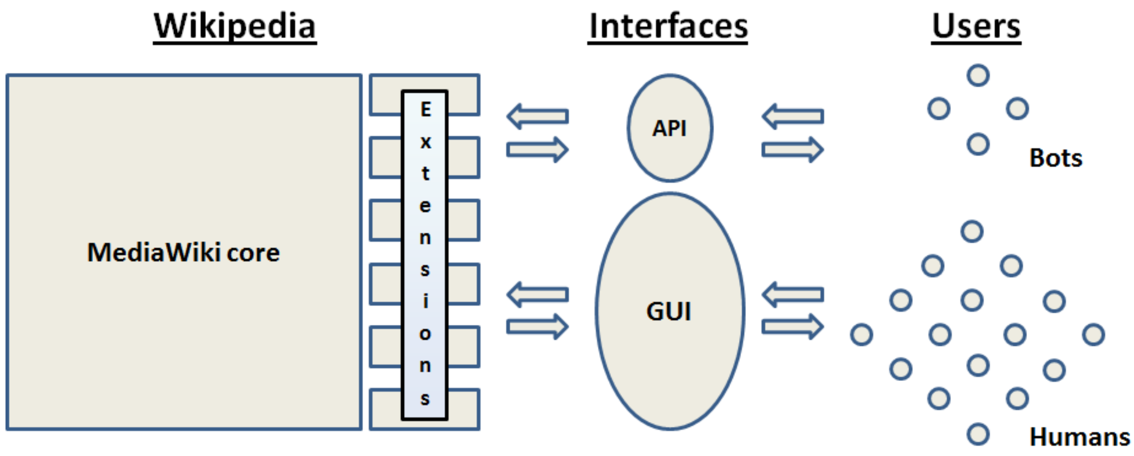

Figure 22. Wikipedia / user interaction flowchart.

The Wiki

As previously discussed, Cunningham developed the wiki with simplicity in mind. Leuf and Cunningham (2001) summarize the essence of the wiki with the following statements:

- A wiki invites all users to edit any page or to create new pages within the wiki Web site, using only a plain-vanilla Web browser without any extra add-ons.
- Wiki promotes meaningful topic associations between different pages by making page link creation almost intuitively easy and by showing whether an intended target page exists or not.
- A wiki is not a carefully crafted site for casual visitors. Instead, it seeks to involve the visitor in an ongoing process of creation and collaboration that constantly changes the Web site landscape. (p. 16)

These statements coalesce around the idea that a wiki is easy to use, a trait that is reflected in its interface, its manipulation, and its online structure.

As users of Wikipedia quickly realize, wiki pages are at heart aesthetically plain—a white backdrop with standard black font and blue hyperlinks. Tables of content, info-boxes, section subheadings, templated messages, and citations are other common features of a well-developed article, while photos, images, figures, videos, sound clips, and other media files can appear as content as well. Even with these added elements, though, wiki pages appear rather conservative compared to most Web content. The appearance of Wikipedia can be customized by users, who can apply skins and personalized CSS scripts, but these will not alter the fundamental structure and layout of pages, which always default to the wiki philosophy of simple display and easy navigation.

Another fundamental element of the wiki page aesthetic and functionality is the high concentration of hyperlinks. Pages should be thoroughly *wikified*,[41] or written with the wiki philosophy that all content that can be linked to other wiki pages should be linked. Wikipedia users will also sometimes come across red links in an article; red links are used to indicate a page does not yet exist for that content, but the author believes it is noteworthy enough for a page, and future contributors should take up the task of writing the new content. The idea of linking to as-yet uncreated content is original to Cunningham's concept of the wiki, as it helps create a site in perpetual expansion, unbounded by traditional standards of inclusion. On Wikipedia, red links "encourage new contributors in useful directions, and remind us that Wikipedia is far from finished"[42] (Wikipedia, 2012cc).

Simplicity is also at the heart of editing a wiki. Instead of dealing directly with HTML code, the basic language of the Web, wikis employ a simplified markup syntax that is closer to natural language writing. For example, instead of a set of HTML tags to

---

[41] *Wikify* is also used more generally to represent a number of layout and formatting conventions of a wiki page.

[42] There are movements within the Wikipedia community to use WikiProjects and social spaces to keep track of unwritten articles in lieu of red links, keeping articles more clean and crisp.

indicate a hyperlink in the text, wikis usually utilize CamelCase or Free Links[43] to designate a link. Wiki markup is not universal to all wikis, but the spirit of straightforwardness is; such a system is meant to overcome the barrier to participation that more advanced programming skills often create. Web browsers, however, speak the language of HTML, so wiki markup relies on the wiki software to translate the code, thus moving this traditionally client-side responsibility to the server-side.

This latter point should not be underestimated. Not only is most Web programming centered on the use of HTML or even more advanced languages, but Web content creation usually necessitates a client-side program installed on the user's computer. The wiki engine, in line with many Web 2.0 platforms, eliminates this requirement, as users only need a Web browser to contribute, again reducing barriers to entry and empowering the user to spend more time and effort on content creation than on technical formatting. The usability of the wiki is certainly not issue-free (the sociotechnical implications of its interface will be discussed later in this chapter), but it does demonstrate the consistent approach to accessibility offered by the technology.

MediaWiki

Although the wiki embodies many of the key principles of new media platforms in the interactive Web era, it is hard to argue this technology would become so iconic had it not been adopted in the early days of Wales and Sanger's encyclopedic experiment. Wikipedia has perhaps done more for the wiki than the wiki has done for Wikipedia, but in either case, the two have become fundamentally linked. There are a number of free and

[43] CamelCase and Free Links are described in Chapter IV.

open-source wiki software packages[44] available for public or private installations, but it is MediaWiki, the engine that drives Wikipedia, that has become the standard bearer for the concept, the philosophy, and the technology.

In the early days of the project, though, both Wikipedia and MediaWiki were still finding their footing. By the time the WMF was founded in June of 2003, Wikipedia was inching into the spotlight, with growing coverage in the mass media, blogosphere, and Web community. Increased exposure brought increased traffic to the sites, and daily page requests grew by an order of magnitude in less than a year (Figure 23). At this time, the English WP alone had passed 100,000 articles, and by the end of 2004, the project contained over 1 million articles in over 100 languages (Wikipedia, 2012d). With the increased activity both by editors and general readers, Wikipedia's performance suffered, with the site frequently slowing down or becoming unavailable for viewing or editing. Developers continued to tinker with MediaWiki's code to improve performance, but they believed the core software still could scale with the increasing size of the project, so other ideas were explored to improve the site's performance.

A solution to Wikipedia's performance problems proved to be a combination of software and hardware enhancements. Until June of 2003, Wikipedia was run from a single computer that acted as both Web server and database storage. A second server was added that month to act as a database server and load-balance resources, freeing the Web server to be more responsive to page requests (MediaWiki, 2012d). Nine additional Web servers were added in January of 2004, and over thirty more by the end of the year, which improved response times to Web visitors. But database responsibilities were still assigned

[44] DokuWiki, JAMWiki, PhpWiki, and Zwiki, to name a few.

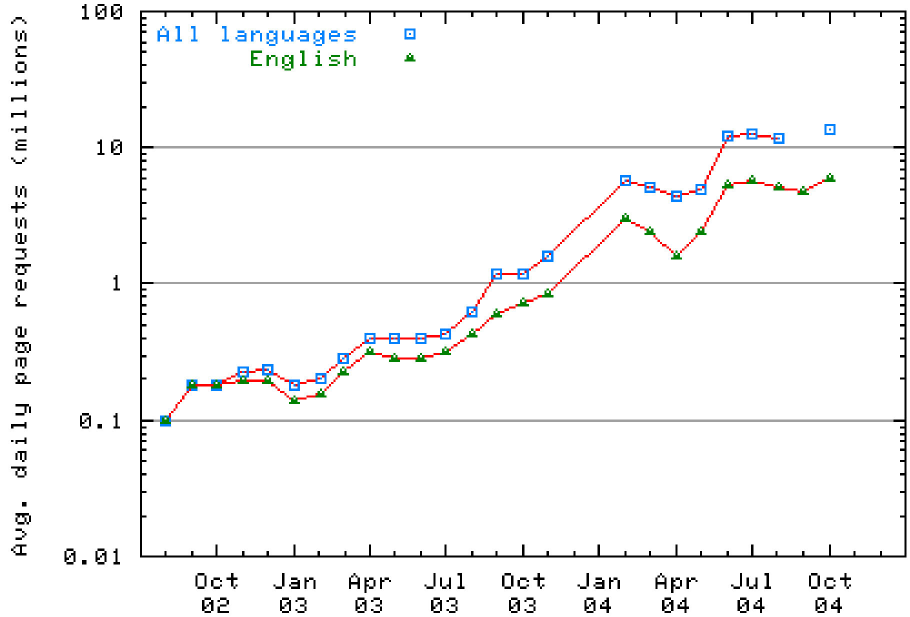

Figure 23. Daily page requests for Wikipedia.org. Chart from User:Kbh3rd.

to one machine, creating a bottleneck of information requests that continued to slow the system (Lih, 2009). To truly improve performance, developers needed to consider the nature of database requests and figure out how simple requests (i.e. displaying a page to be read) and more resource-intensive requests (i.e. editing a page) could be managed more efficiently. The solution was a reorganization of servers and the use of caching.

Developers realized that most Wikipedia page requests were simple requests that did not need to burden the main database, but instead could be handled by servers that hold in memory popular and recent pages that do not change often. The Wikimedia development team turned to a squid server set-up using memcached (Lih, 2009). Squid is "a caching proxy [that] reduces bandwidth and improves response times by caching and reusing frequently-requested pages" (squid-cache.org, 2012), while memcached[45] is a "high-performance, distributed memory object caching system … intended for use in speeding up dynamic Web applications by alleviating database load" (memcached, 2012). Memcached facilitated the memory and performance of the three squid servers set up for simple Wikipedia page requests, keeping this load away from the main database. If the squid servers could not respond to a request, the main database would be queried. The results of the new set-up were "dramatic" and quadrupled the capacity of the Web servers to quickly respond to page requests (Lih, 2009, p. 78). Backup database servers were added as well to increase redundancy (Figure 24).

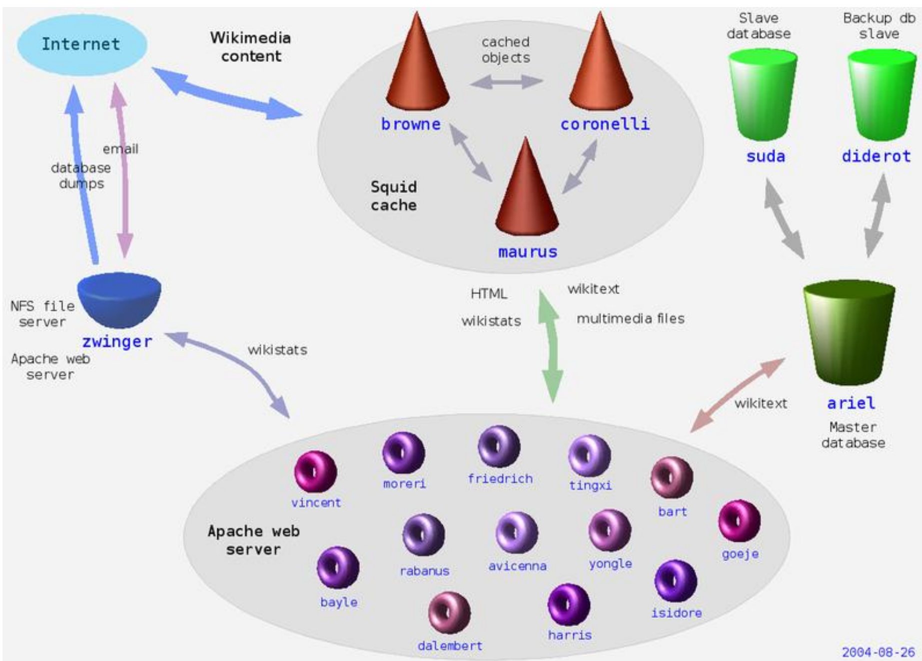

Figure 24. Wikimedia server configuration, 2004. Image from User:Marco Krohn, Wikimedia Commons.

[45] Memcached is an open-source solution, originally developed for the online blogging and journaling site LiveJournal.

As Wikimedia content, contributors, and (perhaps most pivotally) funding have grown dramatically since 2004, so has the size and complexity of Wikimedia's software and hardware infrastructure. Much of the core architecture remains similar to the caching system originally implemented to improve the sites' performance, however. Again relying on open-source solutions, Wikimedia uses a LAMP software bundle to implement MediaWiki, manage databases, balance traffic, and return results for Web requests (Figure 25). LAMP stands for Linux, Apache HTTP Server, MySQL, and PHP, a group of software packages that shares a "development philosophy and tools sets" and has "become popular because it is free of cost, open-source, and therefore easily adaptable" (Wikipedia, 2012e). Wikimedia developers describe the current setup as "LAMP … on steroids," as added redundancies, external storage, and search engines have been integrated into the architecture (Kattouw, 2011).

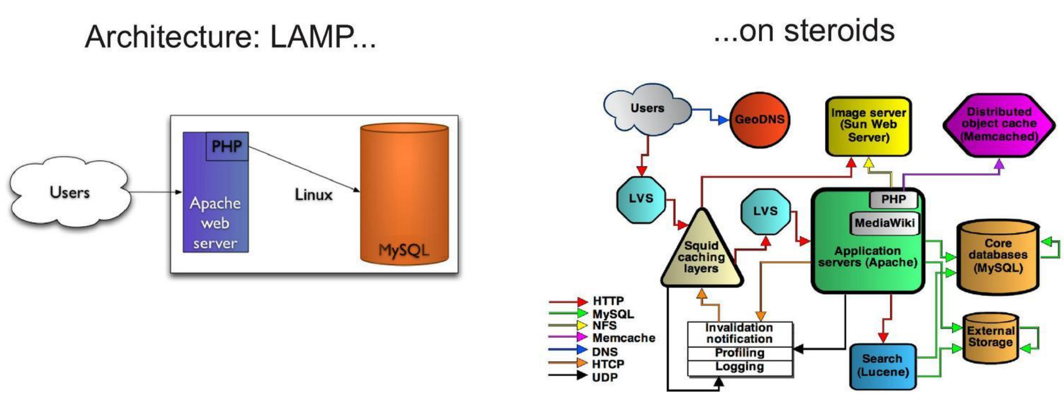

Figure 25. Wikipedia's LAMP architecture. Images from Roan Kattouw, Wikitech.

MediaWiki itself is free and open-source, but its core was developed and has been expanded specifically to run Wikimedia sites. As such, the WMF takes the lead on MediaWiki's development and updates, although volunteer developers still play a role in its evolution. Developers are clear in the online manual about what the software is and is not: MediaWiki is free, server-side wiki software "geared towards the needs of the Wikimedia Foundation" where it is run with a large hardware infrastructure (MediaWiki, 2012c). As such, MediaWiki is not designed for projects needing restricted access and tight security measures, and the software may not be appropriate for smaller websites with limited resources.

MediaWiki is written in PHP, a popular, freely licensed programming language. PHP enables server-side scripting and is used primarily by websites that access and display server-side data. To explicate, consider when a user pulls up a Wikipedia article in her browser. The displayed article is not being accessed from her computer's hard drive (client-side), but rather MediaWiki is interpreting her request for the page, searching for and retrieving the data from one of the WMF's datacenter servers, and sending the data back to her browser to be displayed. All of this is facilitated by carefully crafted PHP code embedded in the article and interpreted by the server. In addition, PHP works with MySQL queries to write and read to databases.

Of course, MediaWiki does more than merely return search results; the software's design and interface are what make collaboration on Wikipedia possible. MediaWiki content—the actual articles and discussion pages, etc.—is written using a "lightweight wiki markup" syntax that is more intuitive than HTML and features simple linking to wikify content. The software was written to handle simultaneous editing and can negotiate edit conflicts when two users are modifying the same content at the same time. Edits are immediately posted to the Recent Changes page, where anyone can watch the real-time changes made to a project. MediaWiki allows the editing of page subsections, making it easier and quicker for the contributor to locate their area of interest. And since MediaWiki was built for all of the WMF's global projects, its interface can appear in a number of languages.

MediaWiki handles a broad range of content types, but uses a uniform page structure to organize and display content. The software supports rich content, which means it can store and display multimedia files like photographs, videos, and animations. It can also support specialized fonts and formatting for content like mathematical equations, musical notation, and Egyptian hieroglyphs (Wikipedia, 2012f). And as explained in Chapter IV, each page uses a tab convention near the top to separate main article content from its associated discussion. A key feature adopted in the earliest versions of MediaWiki and a departure from Cunningham's idea of all content on one page, closely associated but separate article and Talk content is a logical and elegant way to maintain both functionally readable content and ongoing discussions around that content. This set-up has become a hallmark of Wikipedia.

Extensions

A consequence of MediaWiki remaining light and flexible is that many useful features and functionalities are not integrated into the source code. Instead, MediaWiki supports extensions and gadgets developed by programmers to customize their wiki experience. Extensions are scripts run on top of MediaWiki that can be used to extend the functionality of wiki markup, add new reporting and administrative capabilities, change the look and feel of MediaWiki, and enhance security (MediaWiki, 2012b). Both WMF developers and private, third-party developers work on extensions, with WMF Volunteer Development Coordinator Sumana Harihareswara stating the latter group is often motivated by a wish to enhance their private or work-related MediaWiki installations. A gadget is a smaller type of extension, a "snippet" of code that is run from the MediaWiki namespace, and thus is only developed by MediaWiki administrators. According to MediaWiki developers, "This is as it should be: only users especially trusted by the wiki community should be able to edit JavaScript code that is used by other users, since JavaScript can easily be used to hijack accounts or spy on people" (MediaWiki, 2012a).

Extensions have been supported by MediaWiki since version 1.2 was released in 2004, and today there are more than 1,800 extensions available to customize the software. Extensions certainly play a major role in understanding the wiki as a new media object and Wikipedia as a sociotechnical system, but other types of software tools outside of MediaWiki have a more significant connection to this project: semi-automated tools and bots.

Semi-automated Tools

Outside of MediaWiki's core functionality and the features that can be added through extensions, many user-created tools exist to augment the Wikipedia experience in a number of ways. Browsing, searching, editing, downloading, importing, and exporting can all be made easier and more efficient through the use of tools, which can be in the form of short user scripts executed on a page, add-ons for a preferred browser, or stand-alone programs that works with Wikipedia content. Many of these tools are also hosted on the Toolserver[46] and can be run using the Toolserver's database access and processing power.

[46] The Toolserver is a cluster of computers operated by Wikimedia Deutschland that "hosts various software tools written and used by Wikimedia editors" (Wikimedia Meta-Wiki, 2012f). The Toolserver's resources offer computing power to handle large and recurring tasks, and the Toolserver maintains replications of all language Wikipedias.

Of specific interest for this research are semi-automated tools that assist contributors with repetitive editing or maintenance tasks, as these tools are particularly situated at the intersection of social and technical behavior. The key element that makes a tool semi-automated is the requirement that each edit be reviewed by a human user before being committed to the project. These tools will often find new content or questionable situations, including possible vandalism, that require human consideration to decide if intervention is necessary. Assisted editing tools "help power users get dull tasks done quickly (though editors are always responsible for the edits they make, regardless of whether they used an automated tool or not)" (Ayers, Matthews, & Yates, 2008, p. 210).

Some of the most popular semi-automated, assisted editing tools are AutoWikiBrowser (AWB), Huggle, and Twinkle. AWB is a Windows-based browser and text editor that suggests and facilitates mass formatting and housekeeping work, such as standardizing headers, repairing bad links, and correcting typos, capitalization, and duplicate word mistakes. More than a rarely used tool, though, User Magioladitis, one of AWB's developers, points out that nearly 25% of all edits made to the English WP were completed with the help of AWB. Many interviewees also reported that AWB serves as a gateway tool to becoming more involved in the bot community, as these users eventually look for fully-automated solutions to problems on the site. The tool itself can be modified to run in a fully-automated mode, and some approved bots are AWB-based.

Huggle and Twinkle are semi-automated tools largely used for anti-vandalism work on Wikipedia (Figure 26). Huggle, a Windows application written in Visual Basic .NET, was specifically designed to deal with vandalism by monitoring the Recent Changes feed and enabling quick reversions. The tool is used on many versions of

Wikipedia, though on the English WP, certain administrator-level privileges are required for Huggle to be useful. Twinkle is a JavaScript that provides the user with quick access to a number of maintenance tags, warnings and welcomes, and reporting functions. As a script, Twinkle works directly on a Wikipedia page.

Figure 26. Various userboxes indicating the use of semi-automated tools.

These semi-automated tools can be quite helpful for power users who have experience with large tasks, but they can also be problematic when used incorrectly. Though this type of editing is now addressed by the policies governing fully-automated bots, the use of semi-automated tools remains a grey area that some find even more worrisome than the use of bots, precisely because oversight is lacking. Numerous interviewees expressed frustration that the misuse of semi-automated tools like AWB actually creates a bad impression amongst the general community towards any type of script-assisted editing, including bot work.

Bots

> A bot (derived from 'robot') is an automated or semi-automated tool that carries out reparative and mundane tasks in order to maintain the 3,924,146 articles of the English Wikipedia. Bots are able to make edits very rapidly and can disrupt Wikipedia if they are incorrectly designed or operated. For these reasons a bot policy has been developed. (Wikipedia, 2012t)

In many ways, it is hard to determine whether the preceding statements are a definition or a warning, but these lines open the Wikipedia:Bots page for anyone who stumbles upon it. Bots have had a presence on Wikipedia nearly from its start, and today they are a vital element to maintaining and improving both Wikipedia and other Wikimedia sites; as User Mbisanz claimed, "Given the size of the project, the number of editors, and the size of the Internet, it would be impossible to function without them." Their history, though, has been marked by a number of conflicts and controversies, a fact reflected in the anxiety of those opening lines. The following section provides an overview of what bots are and what they do on Wikipedia.

What Bots Are. Bot policy offers its own definition of a bot, contrasting it to an assisted editing tool:

> **Bots** are generally programs or scripts that make automated edits without the necessity of human decision-making.
>
> **Assisted editing** covers specifically lower-speed tools and scripts that can assist users to make decisions but leave the actual decision up to the user. Any program or tool which does not allow the user to view each edit and give an instruction to make that edit (that is, one which can edit without the operator looking at and approving the change) is considered to be a bot. (Wikipedia, 2012s).

The key distinction here is the level of human intervention; bots are unique in that they are client-side programs with permission to make edits without human oversight of each action. Moreover, since bots have the ability to edit at high speeds, their work is automatically suppressed from Wikipedia's Recent Changes feed, removing a second level of human oversight.[47] Thus, we begin to get a sense of why statements of caution might be built into a definition of *bot*.

Bots are in essence immaterial actors; they are lines of computer code that form a script, or a small program that is easily executable, to carry out an automated task. Perl, PHP, and Python are the most common programming languages used to code bots on Wikipedia, and the MediaWiki site maintains a Python Wikipediabot Framework (*pywikipedia*), a collection of generally trusted tools used to automate work on Wikimedia projects. Anyone can download code from the framework to use "as is" or to modify to meet her needs, and pywikipedia serves as a starting point for many new bot operators wishing to automate a task on Wikipedia. Bots are not part of MediaWiki's code, however, and do not work in conjunction with MediaWiki to increase the core software's functionality, as Extensions do. Bots are separate programmed entities, often interacting with MediaWiki in a similar manner to human editors.

Functionally, bots represent more than lines of code, however, which is why an extensive policy has been written for their use on Wikipedia. A bot is a type of registered account on the site with its own rights and privileges, and in fact, a bot account has a higher access level on Wikipedia than a registered human account.[48] The mere creation of a bot account does not provide authorization to run an automated tool, though. A bot operator must make a formal request via a Bot Request for Approval (BRFA) page, which is reviewed by the Bots Approval Group (BAG), a board of community members

[47] Though bot edits are hidden when first looking at the Recent Changes feed, more recent versions of MediaWiki have integrated the option to show these edits, as well as options to show or hide minor edits, personal edits, anonymous users, and logged-in users. These settings offer Recent Changes patrollers more customized ways to police for suspicious activity on the site.

[48] This is justified by the fact that bots go through a formal review before they are used, and thus, they are deemed "trusted" users.

who oversee the process. If a BRFA receives the go-ahead for a trial period,[49] the page will act as a public forum for comments and concerns regarding the bot's activities. When the trial period ends, the BAG decides if the functionality of the bot is sound and if there is community consensus for the bot's task. Approval by the BAG is not the final step, though, as only bureaucrats have the authority to grant the *bot flag* or *bot bit* to an account; this is the setting that officially designates the account as a bot. A BAG member will solicit a bureaucrat to do this for an approved BRFA. At the time of this writing,[50] there were 700 flagged bot accounts on the English WP (Figure 27) that have made nearly 50 million edits to the project (Figure 28).

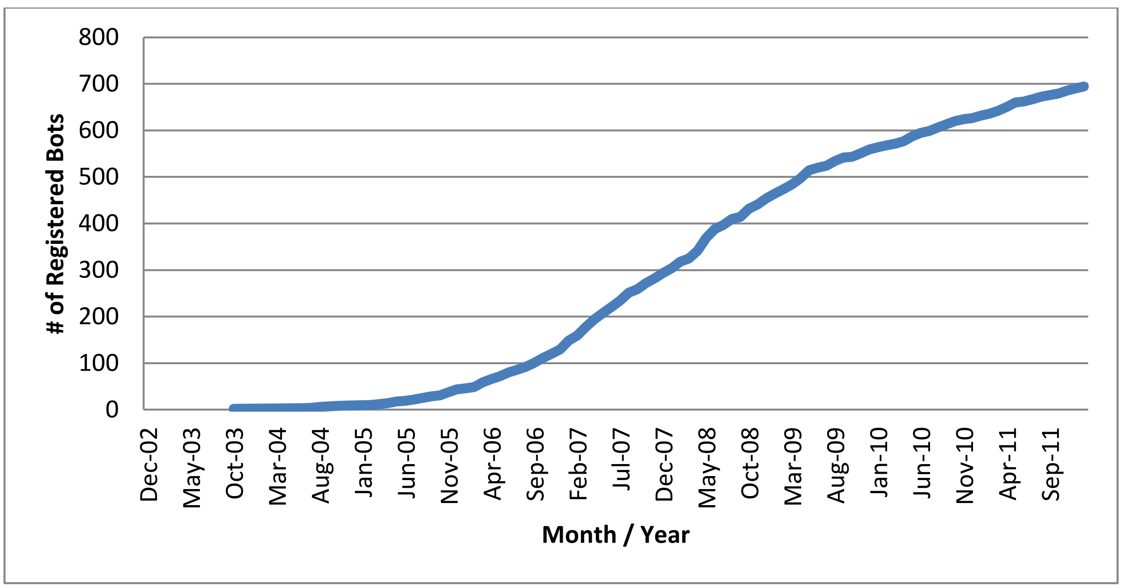

Figure 27. Growth over time of the registered bot population on the English Wikipedia. Data from User Madman.

[49] The usual trial period is one week, though some trials are limited to a certain number of edits. The BAG considers the task the bot will perform when determining an appropriate trial.

[50] As of May 28, 2012. All subsequent bot statistics in this chapter represent data through this date.

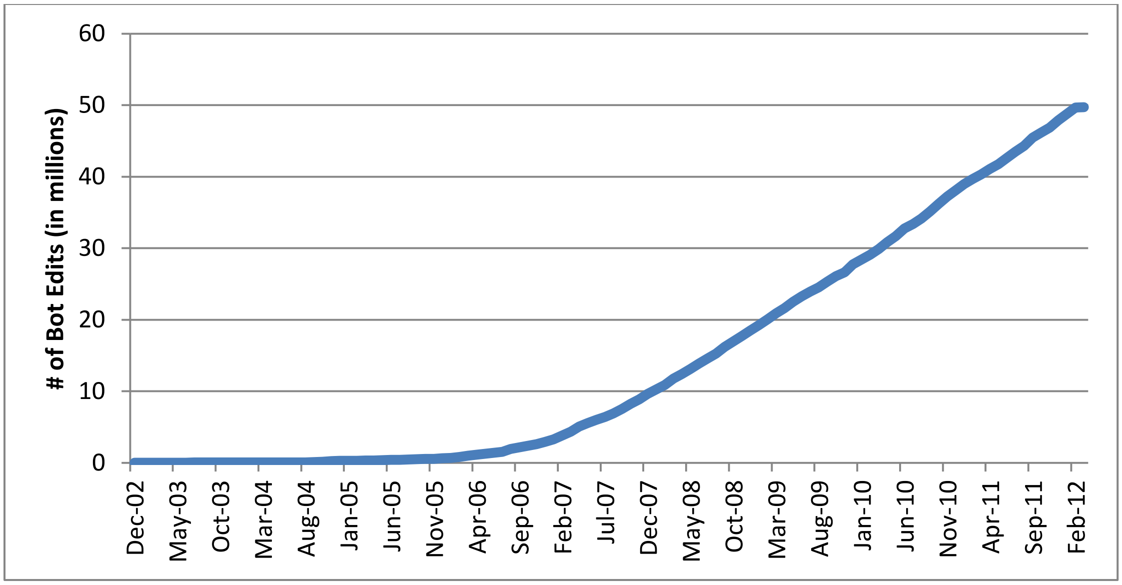

Figure 28. Cumulative number of bot edits over time to the English Wikipedia. Data from User Madman.

A bot account only needs to receive the bot flag when its first BRFA is approved, but many bot operators decide to refine or expand the scope of their scripts, sometimes even tackling brand new jobs with brand new scripts. To do this, operators are required to submit a new BRFA outlining the bot's new functionality, and the request will be approved or denied by the BAG. As a result, there are many more approved *bot tasks* than there are bot accounts; currently, there are 1,554 approved tasks running from those 700 bot accounts on the English WP. A *bot*, then, is generally used to refer to the account rather than the specific scripts run on the account, though prolific bots do indeed build a reputation based on a well-known task (for example, ClueBot's vandalism patrolling).

The user group *bot* is a function of MediaWiki, which manages user rights for all Wikimedia projects, so bots can and do make edits on most versions of Wikipedia (Figure 29). According to many bot operators with experience on numerous projects, the BAG review process for BRFAs on English WP is the most bureaucratic of any language version. On most Wikipedias, a bot task will complete a trial run, soliciting comments from the community, and a bureaucrat will grant the bot flag directly if there is community consensus. Even with a more simplified process, though, the burden of seeking approval on each local Wikipedia for a bot task is quite heavy and discouraging for bot operators who wish to run a task on numerous versions. The global bot flag was established by Wikimedia to alleviate this work by centralizing accountability to one approval process. Once a bot receives a global flag, it is allowed to run on any local Wikipedia or Wikimedia project that recognizes the global bot flag; currently, 397 (46%) of all versions and projects opt-in for global bots. Global bots must have an established track record on numerous wikis and are restricted to running only two types of tasks: maintaining interwiki links and fixing double-redirects. Interwiki links connect versions of the same article across different language versions, thus theoretically strengthening the

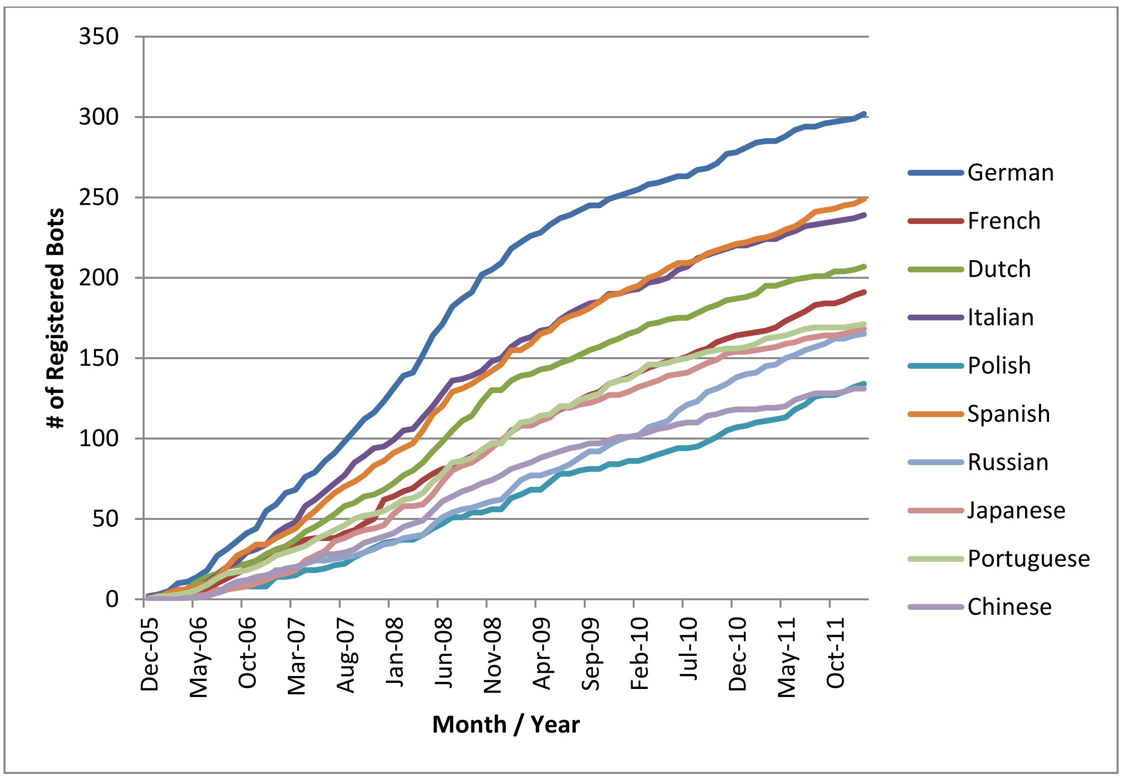

Figure 29. Growth over time of the registered bot population on the ten largest language versions of Wikipedia excluding English. Size is based on number of articles, and the chart's legend lists these versions in descending order. Data from User Madman.

content on each version. Double redirects are a housekeeping failure and usually the result of sloppy page moves; MediaWiki rejects links that will redirect to other pages twice, as they can create an infinite loop. Bots are particularly suited to handle each of these tasks, and global bots are among the most active Wikipedia bots (the most active users of any kind, in fact) as measured by number of total edits.

What Bots Do. The rambot was the first large-scale bot to make an appearance on the Wikipedia scene, adding over 30,000 new city and town articles to the English WP in 2002. These were fundamentally content edits, as the rambot created articles that didn't previously exist and populated them with information pulled from the Web. Since the rambot, other bots have performed tasks to import content to articles, and many bots keep article infoboxes up-to-date.[51] But from the early days of Wikipedia bots, many community members have been particularly suspect of content-editing bots, and a restriction on mass page creation—precisely what the rambot did—has been written into Bot Policy.

Most bots approved today perform administrative tasks rather than content-editing tasks, thus "freeing humans up to do more exciting things" like creating content (User Ucucha). This administrative work includes policing the site for copyright violations, finding and reverting vandalism, blocking spam and spambots,[52] maintaining categories and templates, moving and archiving pages, and producing data reports for administrators. The importance of these bots cannot be understated; they are "absolutely necessary at this stage, in particular anti-vandal bots" (User Rich Farmbrough).

[51] Maintaining current statistics for professional athletes, for example.

[52] Spambots are generally known for sending junk mail to email addresses, but some spambots can also post messages and content to open forums like wikis.

Wikimedia has established a bot policy for bots operating on its projects, including the various language versions of Wikipedia, but most of the sizeable Wikipedias have local policies for their particular communities as well. The two work together like local and national laws; bot operators need to follow local rules and guidelines, though Wikimedia regulations may still be enforced if necessary. As mentioned earlier, the English WP has the most extensive oversight for bots of any WP version, and in turn, the English WP has the most extensive and detailed bot policy as well. According to policy, a bot should demonstrate that it:

- is harmless
- is useful
- does not consume resources unnecessarily
- performs only tasks for which there is consensus
- carefully adheres to relevant policies and guidelines
- uses informative messages, appropriately worded, in any edit summaries or messages left for users (Wikipedia, 2012s)

The policy then goes on to describe each of these points in detail, as well as the BAG/BRFA process and some bot tasks that are generally not allowed. In essence, MediaWiki's bot policy reflects the same principles for the proper use of bots as the English WP policy, though the local policy on the latter is more than four times longer, a sign of perhaps unnecessary bureaucracy that contradicts the community's "avoid instruction creep" ethos (Wikimedia Meta-Wiki, 2012a).

The major restrictions on what bots are allowed to do generally revolve around context-sensitive situations and cosmetic changes. Though there is some debate in the bot community around whether bots can efficiently handle spelling and grammar issues, bot policy prohibits such work from being completed by fully-automated tools, as they cannot sufficiently read contextual clues.[53] Template and categorization work that is context-sensitive, especially around content on living persons, is also prohibited. Exceptions can be made for bots that demonstrate no false positives, but in general, bot operators do not attempt this type of work. Additionally, bots should not make cosmetic changes to the site such as adding or removing spaces or line breaks, or capitalizing words, without strong community consensus or a major revision to Wikipedia's Manual of Style.

Bots and MediaWiki. User Cyde described bots and MediaWiki with the following comparison:

> MediaWiki is big, monolithic, and because it runs the entire site, it needs to be very stable, well-tested, and changes are made slowly. Bots, by contrast, are very light, agile, and programmed and tested at very little risk to the site at large. That's why a lot of tasks end up being handled by bots rather than in the software.

The functional relationship between bots and MediaWiki is actually a complicated one, in ways both beneficial and harmful to each. User Ram-Man pointed out that his early bot efforts helped to improve MediaWiki by identifying bugs in the code, but User MZMcBride worries that some bots can actually harm MediaWiki development by covering up weaknesses:

> There's a bot that goes around renaming categories, which requires editing individual pages to the new category name. If there weren't such a bot, the need for having category redirects (and the ability to move categories easily) would be much more prominent. The same is true with talk page archiving. The bot works so well that people quickly stop noticing how awful talk pages are from a usability perspective.

This research found that communication between bot operators and MediaWiki developers is largely informal and weak at best. Although there is some overlap between

[53] An intentional misspelling, for example.

the MediaWiki development community and the bot community (i.e. some Wikipedians are both), WMF Technical Communications Director Guillaume Paumier stated, “I don’t think there’s a lot of collaboration” between the two groups, and Volunteer Development Coordinator Sumana Harihareswara indicated that the WMF is just now beginning to systematically reach out to the bot community.

Overall, bots have become an integral part of Wikipedia’s larger software infrastructure, though it would be inaccurate to say they work closely with the MediaWiki software the way that extensions do. Largely written by non-MediaWiki developers to tackle tasks, issues, and problems not addressed (or even sometimes created) by the core software, bots are hybrid actors in the Wikipedia ecosystem, made of code and yet positioned client-side with human contributors. To understand the implications of their dual nature, and to better understand this ecosystem of collaboration in general, we need to move beyond functional definitions and explore how the social and the technical fundamentally condition each other on the site.

Wikipedia – A Sociotechnical System

To recall a quote from Chapter II, Bijker and Law (1992) write: “We are witnessing the birth of a new capacity to understand, in a matter-of-fact way, how it is that people and machines work together, how they shape one another, how they hold one another in place” (p. 306). Perhaps beyond what the authors could imagine at the time, technology—especially digital communication technology—has become ubiquitous to our daily lives, and yet at the same time, it is often indiscernible from the social performance of our human experience. Like the air we breathe and vocalize with, the Internet has become a modern necessity for vital information and communication, and yet like air, it is often unnoticed until it is cut off. This contemporary condition—a technological life—makes it all the more crucial to interrogate our relation to the "machines" in our lives along the lines that Bijker and Law suggest.

The preceding descriptions of the social and technical infrastructures that breathe life into Wikipedia attempted to isolate components along a traditional divide, but that divide is in many ways a construction of convenience. Any contributor to the site quickly realizes that Wikipedia is a system of moving parts, an ensemble of structures and agents that are in ways both social and technical at the same time. The following section of the chapter considers the site, and particularly the work of bots and bot operators, from this insider perspective, investigating the processes and flows that dissolve an artificial border between social and technical. In doing so, we see that a more precise, more nuanced way to understand Wikipedia as a new media platform of information and communication is to apply the sociotechnical lens.

How People and Technology Work Together

In many ways, human contributors and bots are virtually equivalent from the perspective of MediaWiki, the core software that runs Wikipedia. As defined in the previous section, users can be conceptualized as separate entities from the MediaWiki software, whether they are human users or bots (Figure 22). While human users and bots do not access data in the same way—human users edit via the familiar graphical user interface (GUI), while bots are expected to edit via the application programming interface (API)—each group interacts with data from a remote position. MediaWiki "sees" contributors merely as accounts, each with its own predefined set of access rights and privileges, rather than as any form of material being. Recognition of bots was built into the original Phase III code that became MediaWiki proper, and Version 1.5 of the software implemented user groups to handle the organization of system rights; the user group *bot* is written into the code along with other user accounts (Figure 30). Though controversial at numerous points in the history of the project, some bots can now also hold administrator rights, aligning them even more closely with a more powerful "human" user group.

```
+$wgGroupPermissions = array(
+       '*'          => array( 'read', 'createaccount' ),
+       'user'       => array( 'read', 'move' ),
+
+       'bot'        => array( 'bot' ),
+       'sysop'      => array( 'createaccount', 'patrol', 'protect', 'delete',
+                               'rollback', 'block', 'editinterface' ),
+       'bureaucrat' => array( 'userrights' ),
+       'steward'    => array( 'makesysop' ), # technically this is for an extension...
+       'developer'  => array( 'siteadmin' ),
+);
```
Figure 30. User groups in MediaWiki (version 1.5) code.

Bots are similar to human users in a number of other ways. Bots receive the same user namespace as any other account, including a user page and a Talk page. Some operators use this space to anthropomorphize their bots, assigning them a gender, posting images or photos, detailing their work, and even writing first-person accounts (Figure 31). Some bots are even gendered by their operators. Talk pages are often where users will ask questions or complain about an action taken by a bot, and bot operators are tasked with watching and responding to these comments, but as User Noommos noted, "I've often encountered users trying to talk to a bot." Bot policy strongly recommends that the word "bot" be incorporated into the name of the account, and bot user pages display a templated message at the top stating "This user account is a bot operated by

______”. Still, bots swim in the same pool as millions of other users, and they are sometimes mistaken for flesh and blood[54] (Figure 32).

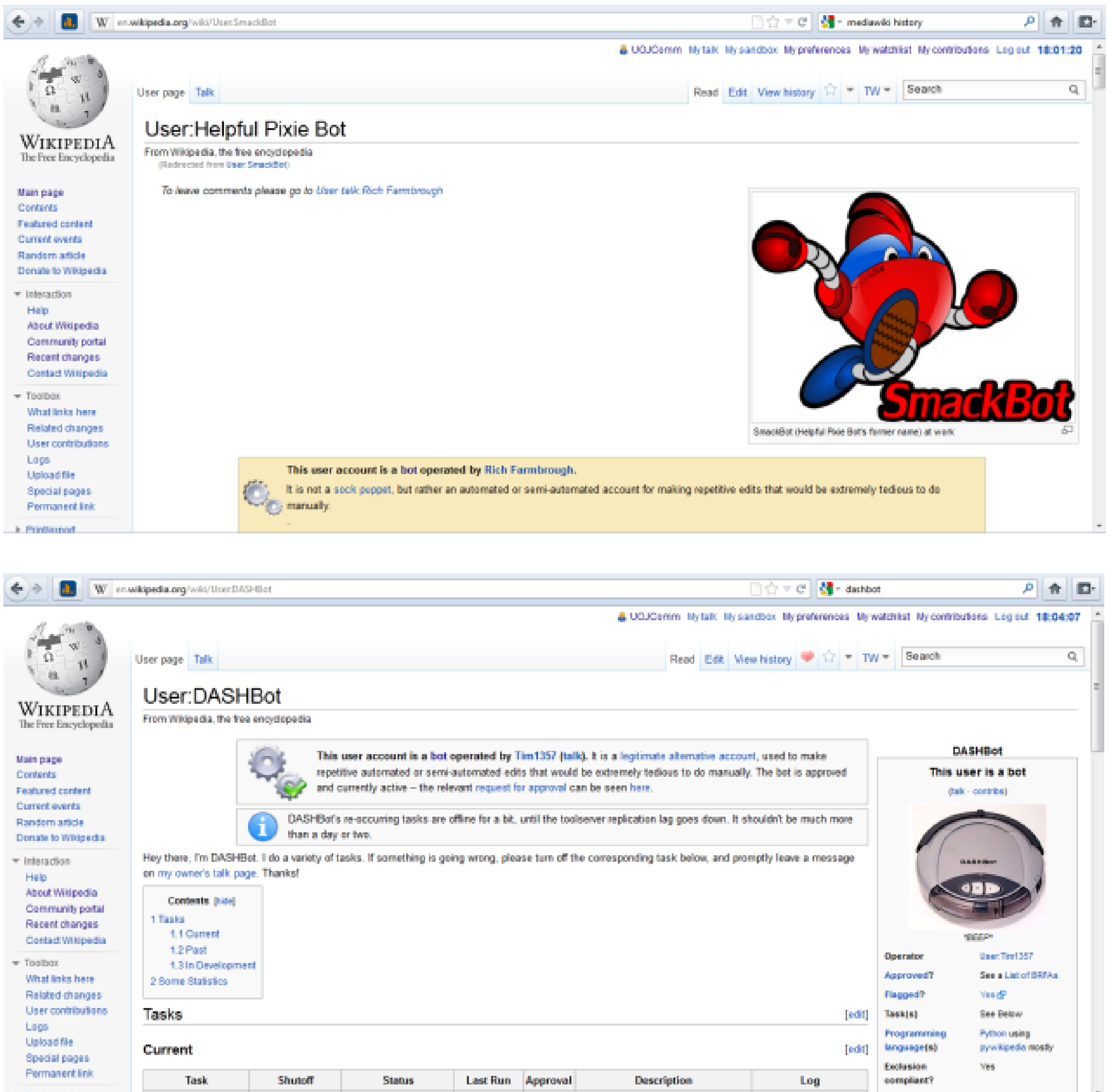

Figure 31. Bot user pages. Helpful Pixie Bot (formerly known as SmackBot) (top), DASHBot (bottom).

[54] Even this researcher has made such a mistake. See Figure 32.

Your recent edits [edit]

Hello. In case you didn't know, when you add content to talk pages and Wikipedia pages that have open discussion, you should sign your posts by typing four tildes (~~~~) at the end of your comment. You could also click on the signature button located above the edit window. This will automatically insert a signature with your username or IP address and the time you posted the comment. This information is useful because other editors will be able to tell who said what, and when. Thank you. --SineBot (talk) 03:33, 22 June 2011 (UTC)

Thank you for the reminder...totally my bad. I'll go back and sign the posts I just made. Thanks again! UOJComm (talk) 03:35, 22 June 2011 (UTC)

Strike that...I see your bot signed my posts for me. Thanks again. UOJComm (talk) 03:36, 22 June 2011 (UTC)

Figure 32. Talk page message from SineBot.

Communication. From the early days of the project, Wikipedians have organized offline meetups to exchange ideas, discuss wiki and non-wiki work, and really just socialize. In addition, two major conferences—Wikimania and WikiSym—have been held annually at international locations since 2005, again to bring people together around their shared interest in wikis and Wikipedia. The WMF and many Wikipedians very much try to bring the spirit of the digital project into the analog world, but these encounters account for a very small percentage of the communication that takes place around Wikipedia; online communication is fundamental to how people work together on the project, and bots are certainly not left out of these channels.

Community members often talk about communication in terms of *on-wiki* versus *off-wiki*, the former being any open and public communication mediated by a wiki page of the related project, and the latter being communication elsewhere online. In keeping with the spirit of transparency inherent to the wiki, most discussions that will result in a change to a project, especially around policy issues, are encouraged to take place on-wiki, where they will be publicly documented and available. Talk pages are the usual home to such discussions, and the consensus process on Wikipedia is intended to be gauged by on-wiki comments only. Even still, a large amount of communication, especially amongst regular contributors and MediaWiki developers, takes place off-wiki on mailing lists, IRC, or through private IM, email, or Skype.

Though they do not participate in on-wiki or off-wiki discussions per se, bots play an important role for human communication through these channels. On-wiki, bots complete a number of administrative and oversight tasks that help make discussions more meaningful. For example, SineBot and its predecessor HagermanBot amend signatures to unsigned comments[55] on Talk pages, while COIBot detects conflict of interest (COI) situations where users may be violating Wikipedia's COI and neutrality guidelines. The MiszaBot family (MiszaBot I, MiszaBot II, etc.) is quite popular for the automatic archiving of discussions on Talk pages and other namespaces. These bots perform functions that maintain the usefulness of on-wiki space.

Off-wiki, a plethora of bots can be found on IRC channels. Chatterbots are notorious for roaming such chat rooms and interacting with users, sometimes providing useful information and sometimes chatting with no identifiable purpose. As many Wikipedia bot operators spend a significant amount of time on IRC, some bots have been deployed there to keep operators abreast of on-wiki activity. A prime example is BetacommandBot3, which lives in the BAG IRC channel and posts an update anytime a BRFA has been edited. A prominent complaint of bot operators on the English WP is a slow review process; BetacommandBot3 reminds BAG members to continually review BRFA pages and facilitates the timely review of requests.

Testing and Enforcing Bureaucracy. Reagle (2010) points out a simple irony of the wiki software: though wikis do not create unnecessary process themselves, their

---

[55] These bots add the missing signatures whether they were omitted accidentally or intentionally, thus potentially implicating them in the social politics of the site, as Geiger (2011) points out.

ability to facilitate content creation makes them fertile ground for the establishment of policies, as policies are themselves fundamentally content. So despite the early guiding principle to “Ignore all rules”[56] and the decade of ensuing debates about “instruction creep” on the project, Wikipedia has become a rich bureaucracy of rules, regulations, and those who enforce them. As exemplified in some of the bots already mentioned, bots and humans work together on Wikipedia to maintain, enforce, and sometimes challenge this bureaucracy.

The bureaucracy on Wikipedia is a sociotechnical one, sprung from both a user base and a software core. Limitations of MediaWiki are both exploited and defended by users, and automated tools are enlisted to take part in this battle. Perhaps the most infamous challenge to both Wikipedia’s social and technical organization came in 2004 and 2005, when a user known as Willy on Wheels[57] began committing mass vandalism[58] on Wikipedia by moving articles to new pages and amending “on Wheels” to the title: for example, “John Kerry on Wheels”. Called one of the top ten trolls in Internet history (Madanapalle, 2009), Willy on Wheels exploited the fact that MediaWiki gave any user the ability to move pages, using a script (i.e. a bot) to strike quickly and broadly. As only administrators have the power to delete pages, and consequently, fix this type of vandalism, Willy on Wheels was also calling attention to and exploiting the social

---

[56] The current policy reads: “If a rule prevents you from improving or maintaining Wikipedia, ignore it.” However, its original language, written by Lee Daniel Crocker in 2002, packs a bit more punch: “If rules make you nervous and depressed, and not desirous of participating in the Wiki, then ignore them and go about your business” (Wikipedia, 2012z).

[57] A later confession sent to the WikiEN-1 mailing list indicates Willy on Wheels was not one user, but actually six people who “chronically trolled numerous wikis, notably the English version of Wikipedia (Willy on Wheels, 2005).

[58] In addition to moving pages, Willy on Wheels also used little-known templates to vandalize Wikipedia’s Main Page and articles found through the site’s Random article feature.

bureaucracy on the site; Wikipedia had a much small number of administrators at the time, so dealing with this vandalism caused a considerable drain on the site's human resources. User Dcoetzee, who studied the case, speculated that Willy on Wheels was at least partially motivated by a loathing of the power structure on Wikipedia and used both advanced technical and social knowledge of the project to carry out the attacks:

> He is deeply familiar with the software and the culture. He knows rarely-used features, notable members of the community not known outside of the community, and appears to be experimenting with his methods offline. … His references to the people who discuss and block his vandalism show that he watches them and revels in the attention, [and] enjoys taunting them. (Wikipedia, 2012k)

Willy on Wheels' edits were quickly reverted and his accounts indefinitely blocked, and by the end of 2005 the user seemingly ceased his attacks.

Willy on Wheels was both a technical and a social attack, but so too was the community's response. Although MediaWiki's openness enabled the attacks, the software also enabled administrators to see what was going on; basic MediaWiki features like Recent Changes allow human users to detect patterns in edits that may be vandalism, and by archiving all previous versions of a page, the system easily enables reversions back to a stable state. In addition, User Curps launched a blocking bot "as an emergency measure … when Willy on Wheels pagemove vandalism reached a new and much more dangerous level," a controversial step that brought the issue of *adminbots* (bots with administrator rights) into the spotlight (Wikipedia, 2012s). More than just an interesting case in the history of Wikipedia, the Willy on Wheels incident demonstrates the sociotechnical aspects of bureaucracy on the project.

Of course, most of the work the humans and bots carry out jointly on Wikipedia is much more benign than the previous incident, though no less important to maintaining consensus-determined policies and regulations. As User Josh Parris comments, “bots are best at pedantry, which is a necessary evil for the encyclopedia.” Bots carry out a number of broad policing tasks, from the detection and warning of copyright violators, to the detection and warning of Wikipedia Manual of Style wrongdoers, to the detection and warning of offensive username creators. Although automated, these bots are not stand-alone enforcers, but rather the policing partners of humans. DeltaQuadBot, for example, which searches the English WP for possibly inappropriate usernames, reports its results to the Usernames for Administrator Attention page for human review. More than just conveying the potential offender, though, DeltaQuadBot offers a bit of guidance for the admin; a username like “Nazik14” is reported with the note:

> Usernames that promote a controversial or potentially inflammatory point of view may be in violation of the username policy due to creating a hostile editing environment- but some real names contain the string "nazi" - especially names from the Middle East - be careful that this is not the case before blocking.

A decision to block is ultimately made by an administrator, but as this example demonstrates, bots can significantly influence these decisions by offering context-specific information and advice that can inform the eventual verdict.

How People and Technology Shape One Another

User 2Legit2 points out in his short essay “There are no rules” that despite the structures of bureaucracy that can sometimes seem dominant on the site, Wikipedia can more fundamentally be understood as a tension between the technical and the social:

> Debates over what “ignore all rules” means are irrelevant.
> There *are* no rules.

> Wikipedia is what the software will allow.
> Wikipedia is what you can get away with.
> If the MediaWiki code allows it, there is no "law" against it. (Wikipedia, 2009b)

There is an essential connection between the social possibilities of the technology and the technology itself; MediaWiki creates a space for a certain type of cyberculture, shaping that culture, but not fully predetermining the nature of that culture. As Levy (1997) reminds us: "The existence of a technical infrastructure in no way guarantees that only the most positive virtualities will be actualized" (p. 191).

At the same time, social factors play a large role in determining what the technology is and how it will be further developed. When Willy on Wheels caused major disruptions to Wikipedia in 2005, MediaWiki developers activated a feature that prevented new accounts from moving pages, a somewhat un-wiki move, but one deemed necessary for the situation. Other proposed solutions for this vandalism tied technical limitations back to social factors like reputation and previous work on the project (Wikipedia, 2005).

All of this indicates that the people and the technology that constitute Wikipedia shape each other in a fluid, ongoing process. This section will examine that shaping by considering how users, computer code, and automated tools influence each other's development.

MediaWiki, Wiki Markup, and the Power of Code. A key element to fully grasping how the social and the technical influence each other on Wikipedia is an understanding that the system is always in flux. Even the most casual users know that the encyclopedic content on the site changes as contributors add and edit information, but policies and procedures, users and bots, and the core code itself are all in a persistent state of refinement. This is elemental to the wiki itself; Cunningham wanted to create a system that made no assumptions about borders, no assumptions about users, and no assumptions about the "end state" of a project—if there ever really is one. Certainly Wikipedia has stabilized around certain conventions and structures, both social and technical, but it carries forward with it the constant possibility of change. The wiki, and in turn Wikipedia, perhaps embody Manovich's variability principle of new media better than any other object; collectively, users play an active role in the creation, manipulation, organization, and consumption of virtually every element of the site.

The MediaWiki core is a prime example of how influence flows both ways (social → technical, technical → social) in this flexible system. The software is developed by a relatively small, paid WMF staff and a much larger group of volunteers. Each group is guided by somewhat different motivations, though they can all be characterized as a confluence of social and technical considerations. WMF developers primarily keep two things front and center in their work: technical optimization and the WMF's strategic goals. Wikipedia is a top global website, and staff developers need to keep MediaWiki as fast and flexible as possible in order to respond to the nearly 500 million page views per day. At the same time, the WMF has set out to make the Wikipedia editing interface more user friendly, while encouraging innovation amongst its volunteer contributors. Those volunteers often contribute ideas and code that is more aligned with their personal uses and installations of the software, but as part of the open-source community and vested members of the Wikimedia community, the social instinct of developing tools and features that can be adopted or tweaked by the public underlies their work as well. All of these impulses merge in a development process that is both organized and “chaotic” (User MaxSem).

The code that Wikipedia content is written in provides another example of sociotechnical influence. Wiki markup, sometimes called *wiki syntax*, is the editing language used by contributors to create content on Wikipedia. The term is generic to any wiki-based website and original to Cunningham’s design, while at the same time used to reference the specific wiki markup format used by MediaWiki to create pages. Theoretically, wiki markup is employed instead of HTML, the basic building block of most websites, as it is more akin to natural language and more intuitive to learn. As Wikipedia’s Technical FAQ describes: “Wikipedia, and wikis in general, are meant to be edited on the fly. HTML is not easy to use when you simply want to write an article” (Wikipedia, 2012x).

In practice, however, many users do not find wiki markup to be quite so simple. As Wikipedia has grown to millions of users, including a much larger base of non-programmers, the usefulness of wiki markup has been reconsidered by many in the Wikimedia community. User Cyde felt:

> One of the big weaknesses [of Wikipedia’s technical structure] is the wiki markup. It was done ad hoc … It does not have a formally defined grammar, and indeed cannot have a formally defined grammar (as opposed to, say, a proper programming language. … More work should have gone into designing this properly at the very beginning.

With the prevalence of HTML interpreters and advanced GUIs on the Web, many now criticize wiki markup as an arcane system and a barrier to entry for new users, an especially worrisome assessment at a time when Wikipedia is seeing fewer new contributors.

If we pause here, we see that the technology of code, written about extensively by new media theorists like Galloway (2004) and Lessig (2006), can greatly influence the social formations that interact with it. A wiki markup that cuts off possible contributors who lack a sense for programming literacy in turn creates a technocracy of the few and skews the very democratic ideals imbued in the original wiki concept. As Lessig (2006) points out, we must never forget the sociopolitical dimensions of technology, and specifically, the code in which our digital experience is written: “Code codifies values, and yet, oddly, most people speak as if code were just a question of engineering” (p. 78).

The WMF, vested contributors, and perhaps most importantly, large benefactors have recognized the sociotechnical implications of the traditional wiki markup and interface, which has not changed considerably since the site was launched over a decade ago. The organization’s latest Annual Report states, “All of the Foundation’s technology initiatives can be boiled down to one goal—reducing the barriers to sharing knowledge” (Wikimedia Foundation, 2011a). In 2009, the Stanton and Ford Foundations funded the WMF’s Usability Initiative, which studied how everyday users interact with Wikipedia and what they find most challenging about the site (Wikimedia Foundation, 2010a). A number of projects to improve both the reading and editing experience have emerged from the initiative, including the UploadWizard to more easily contribute multimedia files, and the Article Feedback Tool to assess content. Most significant, though, is a push towards a Visual Editor that could reduce the off-putting effects of wiki markup. Expected by the end of 2012, the Visual Editor may be “a sufficiently large change to raise MediaWiki’s version number to MediaWiki 2.0,” highlighting the importance of such a project (MediaWiki, 2012e).

Technical Motivations. While MediaWiki and wiki markup have undeniably shaped the Wikipedia and Wikimedia communities, this study's research revealed other, perhaps less obvious ways that technical factors have shaped social formations. A number of contributors cited the tools that are available to administrators as a primary reason for taking up that position. Large and crucial tasks like vandalism patrolling become much more efficient and effective with the page protection, deletion, and user blocking abilities that go with adminship. One editor even claimed to not have "a particularly high opinion of the admin community," but found that "some of the admin tools would be useful" for his work and eventually accepted a nomination. User rights built into MediaWiki clearly have an effect on the ultimate populations that make up each access level (which will be further discussed later in this chapter).

The bot community is shaped in ways by existing technologies as well. Numerous bot operators begin with semi-automated work on Wikipedia before moving to fully-automated bot work. AutoWikiBrowser (AWB) is used by many devoted contributors "to make tedious repetitive tasks quicker and easier," and AWB seems to serve as a stepping stone to bot work (Wikipedia, 2012p). As User Rich Farmbrough put it, "I was aware of bots and was using AWB to fix things faster. Getting a bot flag was a natural next step. I was pretty much in awe of people who ran 'real bots.'" The use of AWB on the English WP is itself restricted to only approved users, usually those with 500 or more main article namespace edits who apply for registration. One does not need to follow this path to bot operatorship (regular user → AWB user → bot operator), and new contributors can certainly prove themselves bot-flag worthy in other ways, but many users do develop in this fashion. As such, the technology here can be seen as a stepwise vetting agent for who becomes a legitimate part of the technical community.

What a Bot Ought Not to Do. The sociotechnical viewpoint is resistant to both technical and social determinism; instead, it looks for the ways each influences and constitutes the other. As we saw, the MediaWiki software that drives Wikipedia is the result of a dynamic sociotechnical system of development that draws on both the power of people and the power of code. Analyzing the details of MediaWiki development, however, would be a vast enterprise that will be left to future studies. There is another new media artifact that offers a more manageable case study to investigate the confluence of the sociotechnical on Wikipedia: the bot.

Bots are immaterial actors in a digital network of information. They are computer scripts—algorithms—that process predefined tasks in an ordered and formal fashion. And yet, as we have already seen, bots are made tangible and social in many ways: by the names they are given, by the avatars used to represent them, and by the voices they are given to speak with. Bots become symbolic actors with tangible effects for Wikipedia and for the community in which they work. The immaterial, technical rules they are given to operate—the definitions and condition statements, the loops and returns—are not the only rules they must obey. Indeed, what bots should and should not be able to do is both a technical and a social question, as evidenced by the responses of this study's interviewees.

Some bot operators feel any task should be within the purview of bots, as long as they demonstrate a proficiency to do the task correctly. As User Kbdank71 puts it, "if a bot is written well enough, there isn't anything it shouldn't be able to do," to which User

Cyde adds, “Anything you can write a program to do is actually fair game.” These operators are in the minority, however, as most feel there are certain tasks bots should not be enlisted to carry out. Their reasons run the blurred gamut of the sociotechnical.

In the early days of bots on Wikipedia, some in the community were worried that even limited bot runs (let alone continuously running bots) would sap the limited technical resources of the site. MediaWiki’s API was not developed until 2006, so bot activity would hit the Web servers in much the same way human activity did, potentially slowing down response times or even crashing the system. These considerations were written into early bot policy, and although the technical infrastructure of the Wikimedia sites has grown considerably and these concerns have greatly diminished, many such provisions are still on the books. Bots should not “consume resources unnecessarily [or] make unnecessary Web requests,” and bots should “be conservative in their editing speed,” making no more than one edit every five seconds for more urgent tasks (Wikipedia, 2012s). Seemingly a technical provision, bot speed limits are actually tied to more human, social factors as well; bots are instructed to run more slowly during peak times (daytime hours in North America and Europe) and peak days (weekdays), when human traffic to the servers are highest. Bots should also stay away from purely aesthetic changes—“really stupid stuff, stuff like adjusting spacing in articles (spacing that [doesn’t] even display)” according to User Cyde—as those runs unnecessarily tax the system.

Bot policy also bars bots from making context-sensitive edits, including spelling corrections and changes. A much debated task in the Bot Policy Archives, spelling tasks were outlawed early on. As User Ram-Man describes, “They are a very popular concept, but we made sure very early that these bots would be disallowed. They absolutely require a human element." Most bot operators agree, and many echo Ram-Man's sentiment that the nuances of language and context require human intelligence to interpret: "Bots should not be permitted to do tasks that require human discretion" (User Mbisanz), "anything requiring judgment should be done by people" (User Oleg Alexandrov), "it is always better to have a human complete high risk jobs" (User Addshore). Others, though, offer a technical justification: "It all comes down to the false positive rate" (User BotOp1), "there are too many variables" (User SatyrTN), "spelling fixes and minor issues … use way too much of the limited system resources" (User DeltaQuad). In these varying comments, we see how both social and technical considerations converge to create consensus around a policy, and we also see how a concept like judgment is defined by both the social and the technical. User Josh Parris sums up:

> The restrictions [Bot Policy] places on unacceptable tasks are there because it's been demonstrated or clearly decided in the past that those tasks require judgment bots don't possess. Mind you, some humans don't seem to possess it either, but no bot does.

Other tasks that bots are restricted from carrying out, or where their use is debated, seem to be chiefly determined by social factors. Bots, like any other users on Wikipedia, are not to carry out tasks where there is no consensus among the community, and in fact, this is the primary metric that some bot operators feel should guide bot work. As User Smith609 articulates, bots should not do "anything for which there isn't consensus! Otherwise, if a bot can do it, then I don't see why it oughtn't." Some bots do operate in the grey area of weak or undetermined consensus, with welcoming bots and bots that edit user pages representing two examples. These tasks have a personal component, and both are carried out on many language versions of Wikipedia. But "bots are not very personable" (User Smith609), and User Jon Harald Søby is "a bit skeptical of bots that deal with users in an automatic way—if an editor only gets messages from bots, that feels very impersonal and not very encouraging." Some welcoming bots post a templated message to a new user's Talk page with the signature of a human user willing to answer questions, but User Snowolf feels "there's no point in a machine welcoming users, and if we want to welcome users, it should be a person making the decision to do so." The effects of automated welcoming on new user retention is debatable—does it alienate new users, or is any welcome a useful welcome?—but with weak consensus, bots continue to carry out the task. In the same way, many bots edit personal Talk pages, though tags have been developed so users can opt-out of certain bot runs.

As we've seen, a close inspection of Wikipedia reveals its technical and social elements cannot easily be dichotomized, and in fact, to understand how the system evolves, we must examine how influence flows across a continuum of sociotechnical concerns. In the next section, we will see how a sociotechnical analysis reveals the durability that emerges in a continuously updated system.

How People and Technology Hold One Another in Place

The popular dictum that "Wikipedia works in practice, not in theory" captures the fact that even with a completely open system where virtually anyone can change anything at any time, the project still has formed into a solid and lasting information and social space on the Web. One can point to the bureaucracy and social structures that have emerged, or the development decisions that have fortified the technical infrastructure, but as we have seen, these forces constitute each other in dynamic ways. The result of this sociotechnical ensemble of actors and influences is a relatively consistent and durable system, where humans and technologies work together to manage work, identify threats, and maintain standards. This section will look at how these ensembles constitute a system of checks and balances that provides stability to the site.

Code and Access. Although equality is a guiding principle of the wiki, not all users are created the same on Wikipedia. For smaller projects with a limited number of contributors, full access for all can work, but Wikipedia's size and scope have brought it beyond the point where idealism can maintain a viable operation. Structure is necessary, and roles are important.

MediaWiki is set up to recognize individuals as users with certain rights, privileges, and access levels, assigned based on the groups of which they are members (Figure 33 for table of groups). The rights that each group receives are cumulative in a way; registered users have more resources and access than unregistered users, administrators have registered user rights plus additional resources and access, bureaucrats additional resources and access than administrators, and so forth. This seemingly social hierarchy is more accurately sociotechnical because it is built into the code of the system. This is not unique to Wikipedia, or even to digital systems. Think of one's access to rooms in a building. There is both a social element—the building's occupants have decided to grant you access to the roof—and a technical component—they give you the key that unlocks the door to the roof. MediaWiki works in much the same way, with a consensus building process to determine access rights, and a technical process—the granting of the *bit*—to make those rights tangible.

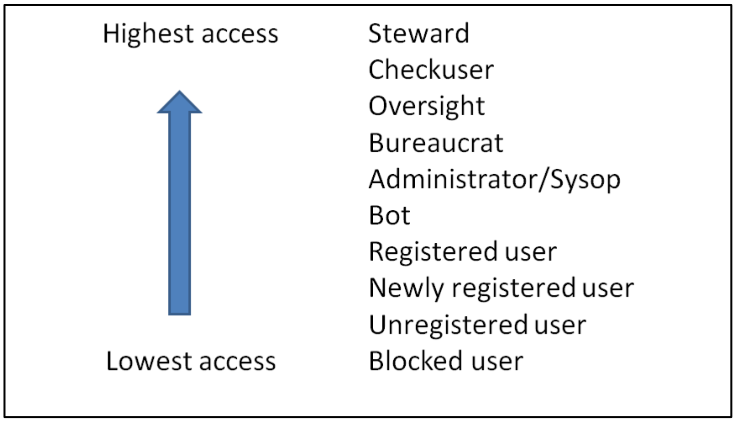

Figure 33. User access rights. Adapted from Niederer & van Dijck (2010) and Wikimedia Meta-Wiki, 2012g.

What is unique to digital systems like MediaWiki is what Galloway (2004) calls protocol, or "a distributed management system that allows control to exist within a heterogeneous material milieu" (p. 7). Galloway (2004) argues that the political economic conditions of modernity have led to distributed, rather than centralized or decentralized, formations of control implemented through technology and best exemplified by the Internet. Protocol is based on "a contradiction between two opposing machinic technologies: One radically distributes control into autonomous locales, and the other focuses control into rigidly defined hierarchies" (p. 50). It is the tension between these two systems from which protocological control emerges, a control that moves from inside out, not outside in. Protocol, or the structured yet flexible schemas that manage distributed technology, "is a solution to the problem of hierarchy," an organization that no longer accurately reflects of our social experience. Protocol is the machine that makes seemingly out-of-control technology "function so flawlessly," the "massive control apparatus that guides distributed networks, creates cultural objects, and engenders life forms" (p. 243). Further, Galloway (2004) argues, "Protocol is a type of controlling logic that operates outside institutional, governmental, and corporate power, although it has important ties to all three" (p. 122).

MediaWiki is protocol for Wikipedia, and the MediaWiki developer community, and in some ways the bot community, are protocological societies. As described in Chapter IV, MediaWiki's evolution has been influenced by the political economy of the WMF, but its logic of organization and control are more precisely explained by the possibilities that are opened up and closed off by its technical structure; protocol deals with the sciences of possibility, not the sciences of meaning (Galloway, 2004, p. 52). User rights built into MediaWiki represent the focused control that competes with the system's distributed "life forms," each with their own locus of power, imbued by wiki philosophy. In this way, we can view MediaWiki as a system of protocol that determines the possibilities for action in the system, though it does not know which possibilities will be realized.

Watching the Watchers. Another way to understand how the sociotechnical elements of Wikipedia keep the system stable is to follow the actors, as actor-network theory prescribes (Latour, 2005). Both core contributors (administrators, power users, etc.) and bots are vested actors on the site who cover a lot of ground (i.e. pages) in their reading and editing travels. Watching is a major task for each group: both watching content and watching other users. In this way, users (be they humans or bots) actually create a network of monitoring that contributes to the health and stability of the site.

Recent Changes and Watchlists are major features of the MediaWiki software that are constantly monitored by regular users. Recent Changes displays the most recent edits to Wikipedia at a near real-time pace. These edits come from all users of the project, and watchers of Recent Changes can both fix honest mistakes as they happen and look for troublesome patterns that may indicate vandalism. Watchlists can serve a similar function, though they are customized by each user to include articles, Talk pages, and other namespaces of interest to that user. These may be pages a user has edited or commented on, but administrators also use Watchlists to surveil pages with a higher likelihood for vandalism and users with a track-record of policy violations. Semi-automated tools like Huggle are also employed to aid in monitoring activities. As vandalism is a constant threat to the site, requiring the consistent attention of those willing to give it, we can say that human users watching other human users is a major task for maintaining the project.

As mentioned earlier, bot edits were suppressed from the Recent Changes feed after early contributors complained they were flooding the channel with unnecessary information.[59] Still, bots are under constant scrutiny from the Wikipedia community, though in different ways at different times. Bots are watched quite closely by the BAG during their trial period, after which they are let loose to perform their approved tasks, be they housekeeping (i.e. changing categories), communicative (i.e. warnings and messages), or authoritative (i.e. reverting edits). At that point, the watching of bots becomes quite reactionary, and many bot operators recounted experiences where they did not learn of a bot malfunction until an innocent bystander brought it to their attention. Incidents of faulty bot work, or even correct work that is disputed, usually face heavy criticism from the community; as User Anomie described, when something goes wrong "they break out the torches and pitchforks." So although in a different way than

[59] Today, one has the option of showing or hiding bot edits in Recent Changes.

human/human surveillance, human users are regularly monitoring and accessing the work of bots.

Many bots, in turn, watch the activity of human users and even other bots, especially around issues of vandalism. Anti-vandalism bots, like the well-known ClueBot and the decommissioned AntiVandalBot, watch users by watching pages. Using blacklists and well-refined algorithms, these bots sniff out suspicious edits, revert those that cross a minimum threshold of potential vandalism, and report the user to the appropriate forums, where administrators can take further action (warnings or blocking). Geiger (2010; 2011) has traced this process in his work on vandal-fighting bots, concluding that “such tools [can] transform the nature of user interaction” (p. 9). Recall too from the Willy on Wheels case that bots can watch other bots as well.

What forms from this network of watching is a complete graph where any user is potentially watching any other user (Figure 34). Some new media critics argue this is the worrisome state of our digital communication systems, one of constant surveillance, and a sign that centralized power structures become only more powerful as they learn from our post-demographics (Fuchs, 2011b; Fuchs, Boersma, Albrechtslund, & Sandoval, 2011). In some ways, the system of watching on Wikipedia proposed here can be exploited to privilege certain users; it has been suggested that administrator nominations are more likely to be approved if the nominee has a strong track record of vandal fighting. However, as watching functions are basic and fundamental building blocks of MediaWiki and available to all users, the control element of surveillance remains largely distributed across the site. Furthermore, as evidenced by the interviews conducted for this study, the Wikipedia community seems to legitimately operate on the founding principle of “Assume Good Faith” and is intrinsically motivated to improve the project rather than wield status or reputation.

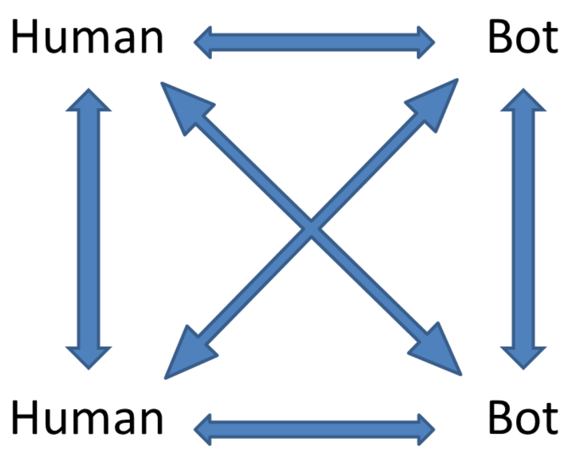

Figure 34. Complete graph representing the ability of any Wikipedia user to watch any other user.

Meatbots. The term *meatbot* is hacker slang for a warm-blooded machine (i.e. a human being). Slightly pejorative, the irony of the moniker comes from the fact that technology is the serious frame of reference here, not humanity. A search for “meatbot” in the policies and guidelines namespace of Wikipedia redirects to a familiar page for this study, Bot Policy, and more specifically, it jumps down to the definition of bot-like editing. The language reads:

> Human editors are expected to pay attention to the edits they make, and ensure that they don’t sacrifice quality in the pursuit of speed or quantity. For the purpose of dispute resolution, it is irrelevant whether high-speed or large-scale edits that involve errors an attentive human would not make are actually being performed by a bot, by a human assisted by a script, or even by a human without any programmatic assistance.

As such, assisted-editing tasks that mimic bot editing in terms of speed or scope should go through the BRFA process like any fully-automated bot task would.

This chapter has argued that to understand Wikipedia as a sociotechnical system, we need to largely reconsider our assumptions about the mythical Wikipedia user. Basic differences and equivalencies between human and non-human users have been discussed, with the suggestion that much of the work done on the site lies in a merged area between the social and the technical. In light of the above policy language, it is worth reiterating here just how much the people and technology that constitute this network of information[60] hold each other in place. "For the purpose of dispute resolution," a very social sphere of the site, meatbots and software bots are the same; really, though, from a broad sociotechnical perspective, meatbots and software bots are quite similar overall.

A Network of Work

Work is a multifaceted concept on Wikipedia that reflects the sociotechnical nature of the system. In 2008, Shirky and Wattenberg estimated that roughly 100 million person-hours of work and thought had gone into the project, but these efforts are divided amongst a plethora of technical, editorial, and social tasks, from writing code to writing articles to writing policies. As we have seen, bots contribute a significant amount of work to Wikipedia as well, but how do we understand bot work in relation to human work? Or more appropriately, how do we understand work that is distributed across an assemblage of actors in a digital system? Looking at material (or in this case, largely immaterial) output is one avenue into this concept, and the following exploratory network analysis looks at a tiny corner of the encyclopedia to begin addressing these questions. Listening to the attitudes and opinions of contributors is certainly another way to understand work, so the results of the network analysis are then discussed within the framework of interviewees' perceptions of human and non-human work on Wikipedia.

[60] Read "information" as *data* and as *meaning*.

Featured Articles represent the best content the English WP has to offer—the best work in a sense—as judged by the encyclopedia's own editors. These articles, then, offer a relevant, self-bounded dataset in which to investigate the network aspects of users and content. Featured Articles go through a nomination, discussion, and consensus building process similar to other promotion processes on Wikipedia.[61] At the time of this study, 3,537 articles (about 0.1% of all articles) held featured status, and these articles are grouped by subject category, ranging from "Art, Architecture, and Archaeology" to "Warfare Biographies." For this study's purposes, particular subject matter is not relevant, though future studies could certainly investigate whether sociotechnical formations differ based on type of content.

Two subject categories were chosen for separate exploration here based on their rich yet manageable size: "Food and Drink" (13 articles) and "Philosophy and Psychology" (12 articles). Data on the total number of edits and total number of editors was aggregated to yield descriptive statistics on the two samples. Data was then converted into a 2-mode matrix, with one mode representing articles and the second mode representing editors. Two bimodal networks (Food and Drink, Philosophy and Psychology) were then constructed and analyzed using the UCINET software.

Descriptive statistics tell one story about the population of editors within each sample (Figure 35). Roughly half of all editors contribute anonymously, with only an IP address identifying their work. Of the remaining contributors, most are registered non-bot users, with bots representing only a small fraction (> 2.5%) of each population. However, when looking at number of edits by group, registered users contribute the bulk of the work, again with bots contributing only a small amount of these edits proportionately (>

[61] For example, Request for Adminship (RFA) or Bot Request for Approval (BRFA).

4%). The high number of anonymous users is not surprising based on previous studies of contributors (Kane, 2009; Viégas, Wattenberg, & Dave, 2004), but these users are more likely to contribute a smaller number of edits (often just a single edit), thus indicating less investment in the articles.

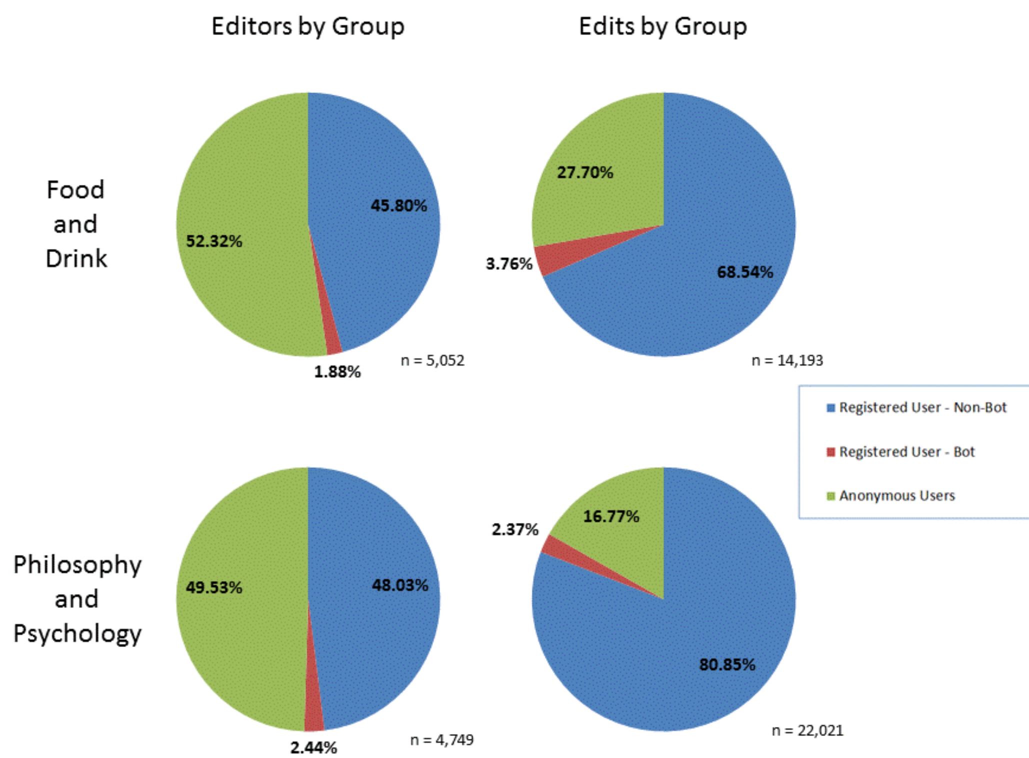

Figure 35. Descriptive statistics on editors and edits by group for the two network populations.

Network measures and visualizations tell a different story about these populations, and particularly about the centrality of bots to each network. Each sample was visualized as a bimodal network using a spring embedding algorithm that groups nodes based on path length (Hanneman & Riddle, 2005). The resulting networks appear as blooms of nodes, with square (blue) nodes representing articles and circular (red) nodes representing editors (Figure 36a). A tie between a square node (article) and a circular node (editor) represents an edit made to that article by that editor. Thus, all relations are representations of work (edits) registered by an editor; direct social relations (editor to editor communication) are not present in these bimodal networks.

The bloom arrangements conceptually make sense with the given data. Around each blue node are congregated the editors who worked on that article. Editors who worked on only one article within a category appear on the perimeter of the network; editors who worked on multiple articles in a category appear more towards the middle of the network, as their multiple ties pull them more to the center. Centrality based on the number of ties to a node is known as degree centrality, and by examining this measure, we can further examine which editors are at the heart of activity in these networks. If we ignore all nodes with degree centrality of one, we eliminate virtually all of the anonymous editors in each population, leaving more dense networks to investigate (Figure 36b). A large number of contributors remain, however, so we can focus in even more by increasing the degree threshold for inclusion in the network's visualization. (Figure 36c)

We can start to see detail on the networks' most central nodes by displaying only actors with a degree of eight or more (Figure 37). Zooming in and applying labels shows us that bots are actually well represented at the heart of these networks. Of the seven most active contributors in the Food and Drink network, three are bots; of the 11 most active contributors in the Philosophy and Psychology network, four are bots. In addition, a number of network measures support the assertion that these bots are central actors for these networks. Whereas degree centrality takes into account the immediate ties an actor has, closeness, betweenness, and Eigenvector centrality are widely used in social network

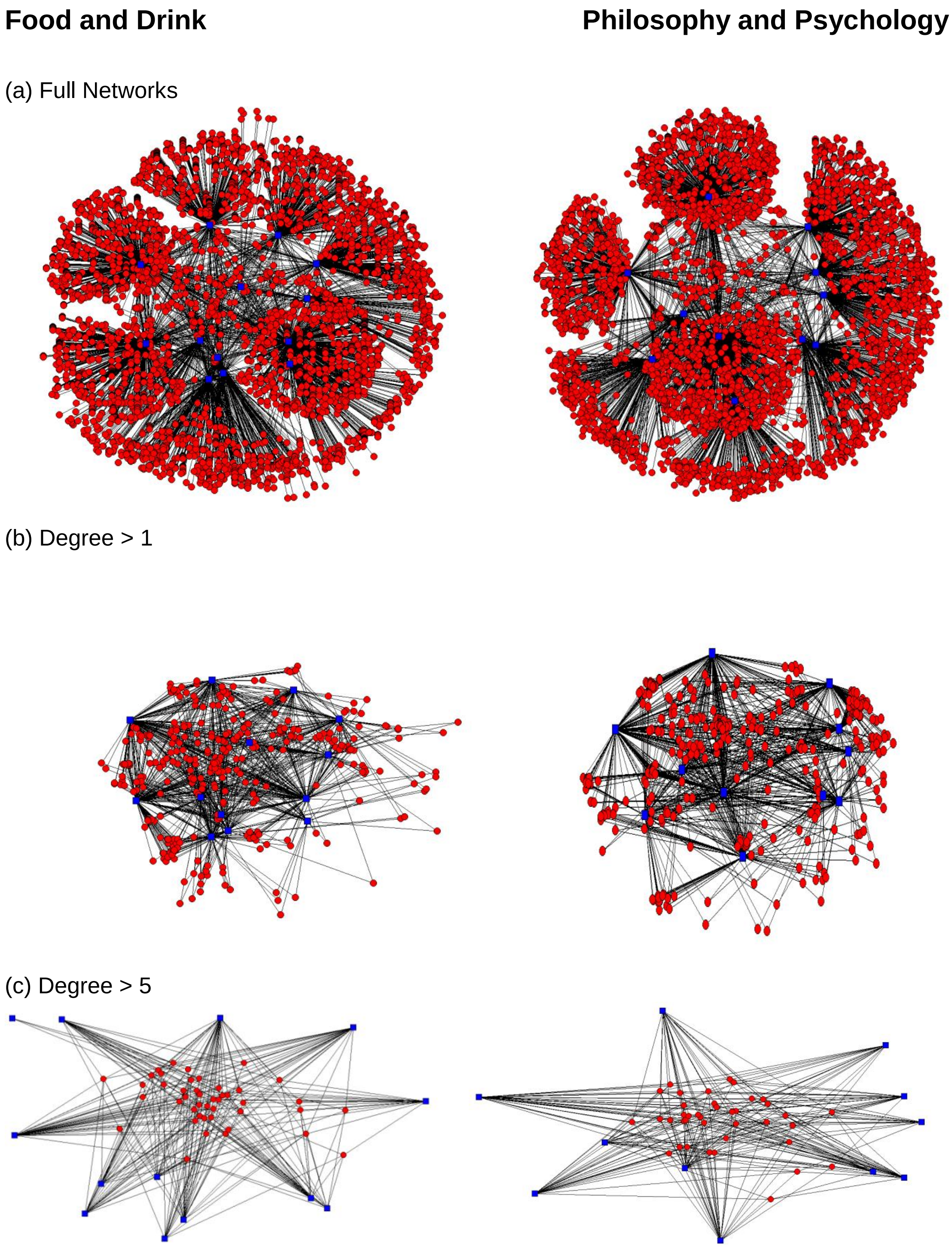

Figure 36. Bimodal network diagrams of decreasing degree centrality. Full networks (a), degree greater than 1 (b), degree greater than 5 (c).

analysis to gauge power at a more global level (Hanneman & Riddle, 2005; Scott, 1991); closeness indicates how close an actor is to all other actors, betweenness indicates which actors are part of the most shortest paths in the network, and Eigenvector centrality uses factor analysis to indicate which actors are tied to the most other powerful actors. When considering these measures, we see that SmackBot is in fact the most central actor in each network, while numerous other bots rank among the most central (Table 3 & Table 4).

What's more, an analysis of user pages and contribution histories reveals that many of the most central non-bot contributors use semi-automated tools to assist their editing at least some of the time. Two of the four highest-degree non-bot users in the Food and Drink network indicate they use such tools (AWB and a user script), while five of the seven in the Philosophy and Psychology network do so, including the now wiki-famous User Koavf. Keep in mind, these are the users who declare their use of semi-automated tools, or use tools such as AWB that indicate themselves in edit summaries; it is likely that other power users employ scripts and tools that are not as transparent.

Analyses of these small networks indicate that the central actors working on these Featured Articles represent sociotechnical ensembles of contributors. Despite their particularly low representation when considering the overall populations in these samples, bots are some of the most central actors in these networks of work. Additionally, many of the presumably human users central to the network are using both widely available and personalized tools to assist their work. It is also noteworthy that some of these human users are bot operators themselves (Brighterorange, Rjwilmsi, Lightmouse, and most notably Rich Farmbrough, who operated SmackBot), thus suggesting that some users find more than one avenue to work on articles. Overall, these

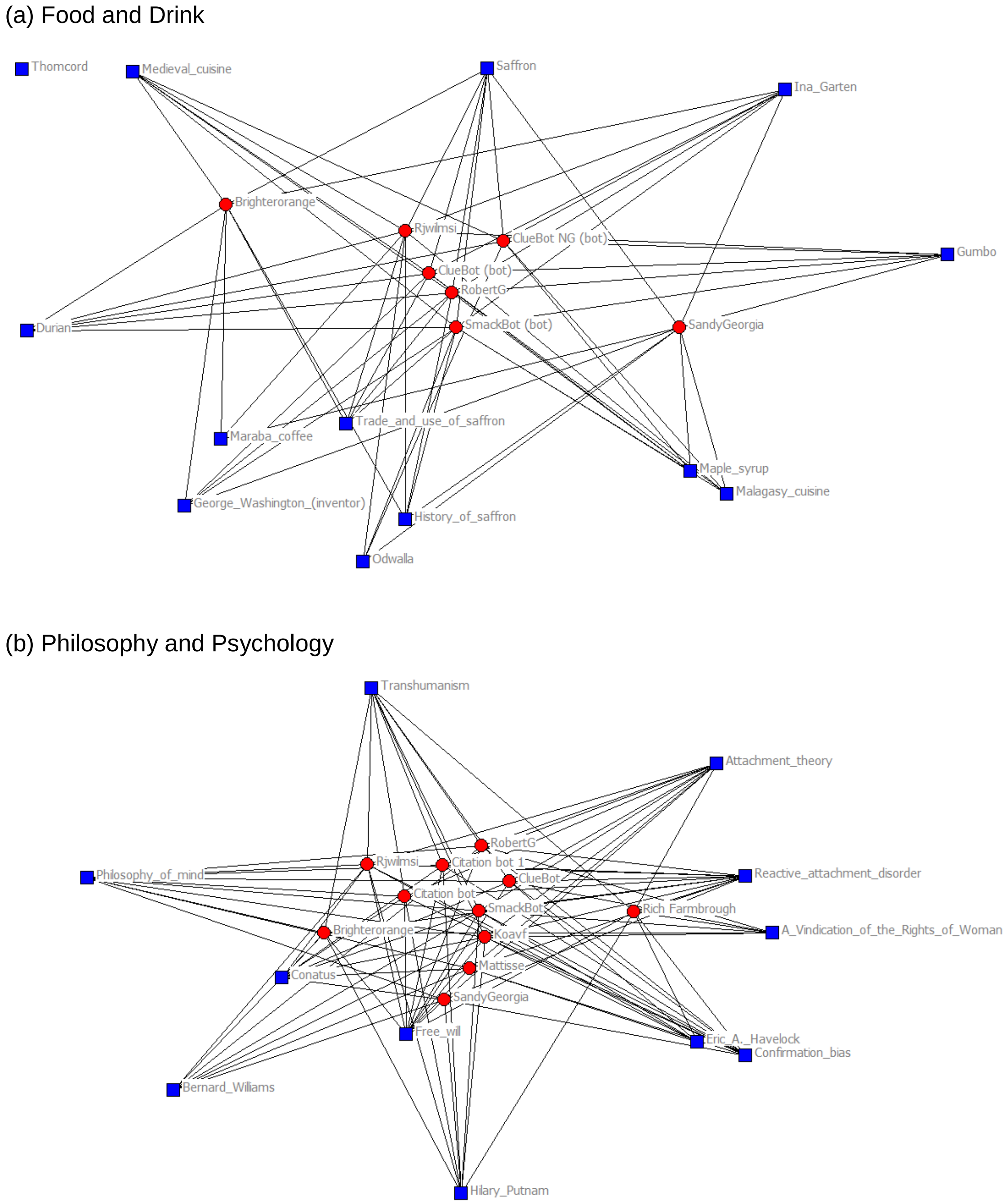

Figure 37. Bimodal network diagrams, degree centrality greater than 8. Food and Drink (a) and Philosophy and Psychology (b).

Table 3

Top 20 Users by Centrality Measures for Food and Drink Network

	Degree		Closeness		Betweenness		Eigenvector	
Editor	Rank	Value	Rank	Value	Rank	Value	Rank	Value
SmackBot	1	0.84615385	1	0.97871309	1	0.03350928	1	0.0561563
Rjwilmsi[a]	2	0.76923078	9	0.80030065	5	0.02160476	108	0.02876313
RobertG	3	0.69230771	2	0.92738610	2	0.02547187	3	0.05391598
ClueBot	3	0.69230771	3	0.92264890	3	0.02535214	2	0.05396639
SandyGeorgia	3	0.69230771	11	0.77682203	6	0.01996121	20	0.04300894
ClueBot NG	6	0.61538463	4	0.88907444	4	0.02233381	4	0.05235607
Brighterorange[a]	6	0.61538463	16	0.74776375	15	0.01229402	1121	0.02594380
AntiVandalBot	8	0.53846156	8	0.80195039	7	0.01697448	11	0.04613167
Can't sleep, clown will eat me	8	0.53846156	12	0.77219635	8	0.01438366	1120	0.02635281
TXiKiBoT	8	0.53846156	17	0.74424248	14	0.01240060	1122	0.02517367
Kwamikagami	8	0.53846156	28	0.70917761	16	0.01191656	1125	0.02298444
Jerem43	8	0.53846156	38	0.68506604	19	0.01053120	41	0.03831227
Lightmouse[a]	8	0.53846156	23	0.72995597	20	0.01038661	1123	0.02469970
Wetman	14	0.46153846	5	0.82753825	9	0.01419951	5	0.04953677
Thijs!bot	14	0.46153846	7	0.80604034	10	0.01412452	7	0.04870262
The Thing That Should Not Be	14	0.46153846	6	0.80991274	11	0.01319358	6	0.04928531
Cydebot	14	0.46153846	10	0.77933585	12	0.01274361	9	0.04695814
Antandrus	14	0.46153846	14	0.75490707	13	0.01272566	27	0.04150295
RussBot	14	0.46153846	20	0.73217517	17	0.01162262	32	0.0400628
MartinBot	14	0.46153846	18	0.74227637	18	0.01156535	29	0.04122829

Note: Bots highlighted in bold. Semi-automated tool users in italics. [a] Also a bot operator.

Table 4

Top 20 Users by Centrality Measures for Philosophy and Psychology Network

	Degree		Closeness		Betweenness		Eigenvector	
Editor	Rank	Value	Rank	Value	Rank	Value	Rank	Value
Koavf	1	1	1	1	1	0.03394026	1	0.05899444
SmackBot	1	1	1	1	1	0.03394026	1	0.05899444
Rjwilmsi[a]	3	0.90909093	3	0.9511755	3	0.02619019	3	0.05803581
ClueBot	4	0.81818181	5	0.90224921	4	0.02506665	27	0.05402939
Citation bot	5	0.72727274	6	0.89041864	7	0.01899605	4	0.05510656
Citation bot 1	5	0.72727274	6	0.89041864	7	0.01899605	4	0.05510656
RobertG	5	0.72727274	4	0.90430897	5	0.02335725	7	0.05396720
Brighterorange[a]	5	0.72727274	9	0.84694880	9	0.01736829	8	0.05383861
Rich Farmbrough[a]	5	0.72727274	10	0.83799028	6	0.02110894	18	0.04922974
SandyGeorgia	5	0.72727274	16	0.78330725	12	0.01448561	47	0.03938593
Mattisse	5	0.72727274	21	0.76208418	13	0.01419763	77	0.03735014
RjwilmsiBot	12	0.63636362	8	0.85795503	11	0.01557969	9	0.05355658
Gregbard	12	0.63636362	12	0.81626170	16	0.01180711	10	0.05310187
Bluebot	12	0.63636362	43	0.69177037	15	0.01201548	1175	0.01833807
Michael Hardy	15	0.54545456	13	0.80383867	18	0.01028743	11	0.05197501
Thijs!bot	15	0.54545456	14	0.79803574	19	0.00981025	12	0.05175314
MisfitToys	15	0.54545456	17	0.78253352	21	0.00903601	13	0.05113545
ClueBot NG	15	0.54545456	11	0.83548641	10	0.01619512	14	0.05042028
Yobot	15	0.54545456	15	0.79258024	14	0.01269825	17	0.04974367
Cydebot	15	0.54545456	23	0.75530308	20	0.00948158	27	0.04688409

Note: Bots highlighted in bold. Semi-automated tool users in italics. [a] Also a bot operator.

sociotechnical arrangements seem to indicate that a relatively strong technical literacy is necessary in order to reach these levels of contribution and commitment.

Zombie Labor

The previous networks indicate the centrality of certain actors (both human and bot) in small networks of Featured Articles, where a relation or tie between an actor and an article represents an edit to that article. Network calculations, then, are made with *the edit* as the basic unit of measure, assuming a certain validity for that unit. Indeed, edit count has played a significant role in assessing the amount of work contributed to the project by a user throughout Wikipedia's history, a fact evident in numerous ways, from the info boxes many users maintain on their user pages (Figure 38) to the fanfare around Koavf's millionth edit. Edit count is described on the project as "a quick and crude aid when trying to measure a Wikipedian's experience in the Wikipedia community" (Wikipedia, 2012v), and many interviewees see its importance as a double-edged sword. User BotOp1 agreed that "it's indicative of [editors'] approximate experience … but it shouldn't be something we base opinions on." User Rcsprinter stated it doesn't matter to him personally, but it does represent his experience. Other opinions are more polarized, usually downplaying the importance of edit count as something that used to matter to them when they began editing, but not something that motivates them any longer. Many cautioned that a preoccupation with edit count can lead to *editcountisis*, prompting a user to make unnecessary edits in order to inflate her count, something that User Oleg Alexandrov suggested experienced editors grow out of.

Figure 38. Userboxes displaying edit count.

Some interviewees offered a more nuanced interpretation of edit count, and work in general on Wikipedia, highlighted by User Rich Farmbrough's assertion that "not all edits are equal." A clear distinction is raised between content work and back-end administrative work. User Kbdank71 explained:

> If I did article content editing, edit count would not matter to me, because it would not represent the sum total of what I did. But not knowing a lot about any one subject … I wasn't interested in article content. I was more interested in making the encyclopedia easier to use; hence [my] category work. So for me, there is a concrete way to determine how much work I did, and that's my edit count.

User Ram-Man used an analogy to describe different kinds of edits: "My edits were cheap and I never made a featured article. I'm like the miner who produces the raw materials while some other architect/artist/builder turns it into something beautiful."

Other contributors draw more of a distinction between the work of humans and the work of bots. User Tedder claims that edit count "doesn't mean anything to me in the context of bots. … It isn't doing real work, it's doing metawork." User Cyde dubs this "zombie labor," explaining:

> The amount of work the bots do, at least on an edit count basis, is very high. In terms of an actual amount of work basis, it's very low. I wrote most of Cydebot many years ago and it's racked up millions of edits since then with very little work from me. … So the average editor who might make a few edits a day is doing a lot more work than I am, even though Cydebot is making tens of thousands of edits a day.

The fact that bots, once programmed, approved, and off running, work with little human effort expended seems obvious—these are automated tools, after all, meant to ease their creator's and the project's workload—but when the end products of their work, the edits, are virtually equivalent to what a human would produce, the concept of work becomes muddled. Remember, bot edits are routinely suppressed from appearing in the Recent Changes feed that displays all non-bot user work on Wikipedia. This arrangement seems to tread perilously close to the system of fetishism described by Marx (1867) as hiding the social relations (i.e. labor or work) behind a commodity, only on Wikipedia, the human relations and contributions are prized, while it is the technology that is subjugated. Wikipedia has built a myth around the power of human collaboration, but as this project has endeavored to show, technology plays a bigger role in the site's story than generally acknowledged.

In the end, though, mixed understandings of what work is and who should do it are often sorted out pragmatically. As User TheFearow stated, "It's simply a matter of labor—you can use one bot, or a few hundred humans." Indeed, the fact that bots seem to be quite central to the development of quality content, even if their work is undervalued and unnoticed, signals a sociotechnical system of resource management that may be needed if the trends indicating contributor decline continue. Negotiations around the meaning of work on Wikipedia may be fluid, be so too is the constant state of the project, and work is always necessary to keep the project afloat.

Conclusion

By digging into the details of the fundamental structures, heterogeneous actors, and everyday activities on the site, this chapter presented the case for a sociotechnical understanding of collaboration on Wikipedia. As discussed, functional distinctions that isolate the project's social and technical infrastructures deny the basic and essential ways in which these structures constitute one another. A sociotechnical approach more properly elucidates this dynamic system, composed of a network of human and non-human actors, technical protocols, and consensus-established policies. Through examining the assemblages of humans and technology at both general (bureaucracy/MediaWiki) and specific (human user/bot) levels of development and interaction, the chapter argues that sociotechnical codependencies drive the maintenance and growth of Wikipedia, providing both necessary fluidity and obduracy to the project. From this understanding, we can more properly examine the site for something even bigger than the sum of its parts: the possibilities for collective intelligence.

CHAPTER VI

WIKIPEDIA AS COLLECTIVE INTELLIGENCE

Psychologist Jean Piaget (1953) once said, “Intelligence is what you use when you don’t know what to do,” referring to the ways that people are able to solve problems despite not always having the best information in their heads or the appropriate experiences under their belts. The concept of intelligence is truly a slippery one, debated through the ages by scientists and philosophers precisely because of its intangible qualities. And yet, to the chagrin of many, IQ tests have become the standard bearer of intelligence in society today; a very inward looking measure of cognitive ability, IQ scores quantify and stratify what it means to be smart (and not so much).

The relatively rapid development of computer technology over the past half century has offered new ways to adapt, revise, and rethink theories of intelligence, though. Work on artificial and machine intelligence in some ways models the standard approaches used to understand human intelligence, and in other ways breaks away from them, exploring the unique attributes of powerful networked systems. With the rise of the Internet and the emergence of interactive Web technologies, ideas on intelligence are becoming more outward looking, going beyond the individual to investigate how groups of actors—be they human or machine, smart or dumb—are building collective intelligence that speaks to Piaget’s point: the proof of intelligence is getting things done when you individually don’t know what to do.

As the previous chapter argued, Wikipedia is best understood as a sociotechnical system where heterogeneous actors and forces constantly condition each other, maintaining a tension that keeps the project stable and yet flexible. An environment like this flourishes from dynamic collaboration, and as this chapter will contend, plays a major role in harnessing the individual intelligence of its actors and enabling the mechanisms for collective intelligence to develop. Ultimately, something greater than the whole—a cyberculture that is concurrently informational, technical, and social—emerges from the sociotechnical system of Wikipedia. To make these arguments, the chapter first reviews theories of intelligence and their applications, illuminating what intelligent activity takes place on the site and how that intelligence is perceived. Specific concepts of collective intelligence are then explored in the context of the site, followed by insights on how this form of CI offers complexities and opportunities beyond more traditional models of intelligence.

Theories of Intelligence

Broadly speaking, both digital/new media theory and STS are concerned with the interactivity of actors in an environment. The frames for these theories are often the form of the environment and the social context in which actors are bound. The study of intelligence, though still contingent on the social and the material, offers an additional viewpoint, one that places the actor front and center. Understanding theories of intelligence—human, artificial, and collective—will provide a more nuanced interpretation of sociotechnical systems of digital/new media, including Wikipedia.

Human Intelligence

Hawkins and Blakeslee (2004) write, "The question of intelligence is the last great terrestrial frontier of science" (p. 1). Despite millennia of philosophizing and over a century of systematic scientific research, human intelligence remains a mystery in many ways—an elusive concept to even define and measure. When Herrnstein and Murray published *The Bell Curve: Intelligence and Class Structure in American Life* in 1994, a work that made overly strong connections between intelligence tests, genetics, and social outcomes, many psychological researchers spoke out about these tenuous claims and sought to reinforce what is known and unknown about human intelligence (Deary, 2001). Representing the larger scientific community's response to Herrnstein and Murray, Neisser et al. (1996) reviewed both past and present research on the topic, arguing that there is in fact a strong consensus about research findings on intelligence, but there remain too many unresolved issues and unanswered questions to make confident claims about the its implications.

The bulk of research on intelligence has used psychometric testing, introduced by Binet and Simon (1980) at the beginning of the twentieth century. A dominantly quantitative approach, this method uses a number of mental tasks to assess thinking, reasoning, and memory, ultimately determining a score for an individual subject. Popularly known as an intelligence quotient (IQ) score, the measure has culturally become a marker for one's level of intelligence relative to her peers. As Neisser et al. (1996) highlight, however, even when IQ scores strongly correlate with other achievements (like accomplishments in school) or traits (like heredity or race), there are too many individual differences and unknown factors for IQ scores to be reliably predictive. In addition, researchers cannot explain the Flynn effect, or the steady worldwide increase in test scores that has seen the mean IQ increase by 15 points over the past 50 years; although improved nutrition, cultural changes in schooling or child-rearing, and experience with testing may be influencing this increase, "no one is sure why these gains are happening or what they mean" (Neisser et al., 1996, p. 97).

As psychometric testing relies on diverse individual tasks including elements of language and vocabulary, shapes and spatial reasoning, numbers and arithmetic, and memory, various models or conceptions of intelligence have been built around the tasks that most strongly correlate to one another. Research and replication have shown that all tasks in an IQ test positively correlate to each other, leading many to subscribe to the *general intelligence* (or *g*) model originally described by Spearman (1927). Deary (2001) describes *g* as "a significant, inescapable fact … that accounts for half of the variability in mental ability in the general population" (p. 13). But as subsets of tasks correlate more strongly, others have suggested hierarchical models with *g* at the top and other significant groupings—such as general memory, cognitive speed, and visual perception—at the next level down (Carrol, 1993). Still others have proposed *multiple intelligence* models that incorporate components not assessed by an IQ test (Gardner, 1983) and cognitive development models that focus on an individual's learning, awareness, and mental development (Piaget, 1970).

Despite these numerous approaches, there is no test that evaluates all forms of human intelligence, and there are still no established measures for many qualities important to the human experience, including creativity, wisdom, and common sense (Deary, 2001; Neisser et al., 1996). Gottfredson (1997), then, offers a definition that takes these dynamics into account:

> Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings—'catching on', 'making sense' of things, or 'figuring out' what to do. (p. 13)

Theories and models of human intelligence have informed much of the work and development of artificial and machine intelligence, and the unresolved questions of the former have pushed the latter in numerous ways as well, to be discussed in the next section.

Artificial and Machine Intelligence

Documented ideas and concepts on artificial life and artificial intelligence (AI) date back to the ancient Greeks and have permeated both philosophy and literature from Homer to Hobbes to Mary Shelley (Lugar, 2005; McCorduck, 2004). More than mere conceptions, though, material attempts to create "automata that move and devices that reason" well predate the computer age. Da Vinci drew schematics for a robotic medieval knight in the fifteenth century, and French inventor Vaucanson built a working mechanical duck that would eat and quack in the eighteenth century (Nilsson, 2010). The computerized attempts to develop AI that emerged in the twentieth century and continue through today draw from a large pool of Western thought.

Definitions of AI are difficult to formulate for two reasons. As previously discussed, a durable and general definition of *intelligence* itself is elusive, as research continues to probe its dynamic meaning for humans. In addition, recent AI research has gone in a number of directions, with both overlapping and differing emphases. Nevertheless, Luger (2005) offers two definitions that seem wide enough to cover the field:

> Artificial intelligence may be defined as the branch of computer science that is concerned with the automation of intelligent behavior. (p. 1)
>
> AI is the study of the mechanisms underlying intelligent behavior through the construction and evaluation of artifacts designed to enact those mechanisms. (p. 825)

The author reminds us in true STS form, however, that "AI, like every science, is a human endeavor, and perhaps, is best understood in that context" (Lugar, 2005, p. 2). The term itself was coined at a 1956 conference at Dartmouth College attended by leading academics and engineers, including Marvin Minsky and Claude Shannon (McCorduck, 2004; Schuster, 2007).

Emphasizing the ambiguous notion of intelligence in both humans and machines, Turing (1950) proposed an empirical test to measure the performance of machine intelligence against the standard of human intelligence. The Turing Test places three actors—the machine, the human, and a human "interrogator"—in their separate rooms, with the interrogator given an equal amount of time to converse with each using a text-only terminal (Figure 39). Based on the conversation, the interrogator is tasked with distinguishing who is the human and who is the computer; if she cannot do so, Turing argued, then the machine may be deemed intelligent. Lugar (2005) argues that the importance of the Turing Test for testing AI lies in its attempt at an objective notion of intelligence, its disinterest in the internal processes of the machine, and its elimination of bias favoring the living human. Although many have criticized aspects of the test, including its limited scope of symbolic problem-solving, it has provided "a basis for many of the schemes actually used to evaluate modern AI program" (Lugar, 2005, p. 14).

Turing was likely influenced by models of behaviorist psychology in the mid-twentieth century, and much of the early computer work on AI focused on machine learning that in many ways mimicked human learning (Hally, 2005). In what Schuster (2007) calls *traditional AI*, intelligence and cognition is investigated from an algorithmic,

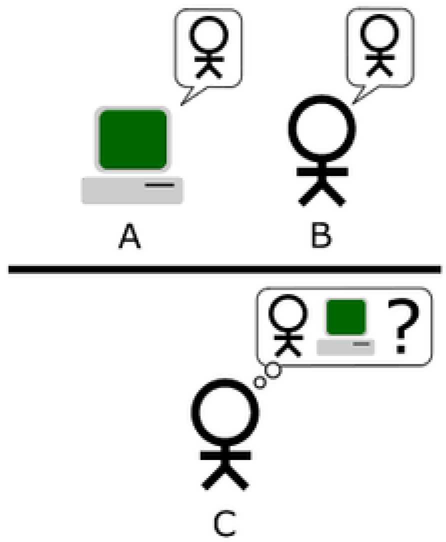

Figure 39. The Turing Test of Intelligence.
Image from User Bilby, Wikimedia Commons.

computational point of view. Machines would use an input-processing-output sequence known as the Information Processing Metaphor for its similarity to the sense-think-act process of the brain (Schuster, 2007). Eliza, the first software robot (bot), developed in 1966 at MIT and capable of carrying on an extended conversation with a human, was programmed this way, using a Rogerian therapy model to take in textual information, reformulate it in an appropriate context, and output it back to its conversation partner (Leonard, 1997). Today, bots compete for the Loebner Prize, given at an annual Turing Test competition of the most sophisticated semantic bots. Despite nearly fifty years of development, though, no bot has won gold or silver prizes for successfully convincing the judge of its humanness (Loebner.net, 2012).

It is significant to note that much of traditional AI research focused on software development, as the size and cost of hardware remained impediments to exploring other aspects of machine intelligence until the 1980s. Since then, the accessibility of cheap and powerful hardware has enabled *new AI* research to investigate "intelligence from the viewpoint of a creative interplay between one or more entities, so-called agents, and a complex, real-world environment" (Schuster, 2007). New AI often employs visual and audio sensors to feed in additional data on a machine's surroundings, as well as robotics to move and respond in the environment. Connecting with theories of embodied cognition from philosophy and psychology, some researchers argue that embodiment is a prerequisite for intelligence and that solely computational models which neglect physical characteristics "are doomed to failure from the very outset" (Pfeifer & Bongard, 2007).

Other recent and emerging areas of AI include ubiquitous computing, bioinformatics and synthetic life, DNA computing, neuroinformatics, quantum computing, and natural language processing (Schuster, 2007). Though traditional, computational AI is most pertinent to the present study, these new areas of inquiry are important for recognizing how ideas on intelligence continue to fluidly evolve and drive technological development contingent on information and communication.

Collective Intelligence

Unlike many traditional conceptions of intelligence that primarily consider cognition for the individual, *collective intelligence* considers the social aspects of intelligence at the group level. For a current, collaboratively-constructed definition of CI, Wikipedia (2012a) offers: "Collective intelligence is a shared or group intelligence that emerges from the collaboration and competition of many individuals and appears in consensus decision making in bacteria, animals, humans and computer networks." The M.I.T. Center for Collective Intelligence (2012b) simplifies this to: "Collective intelligence is groups of individuals doing things collectively that seem intelligent."

These definitions emphasize that intelligence is a quality of actors and systems both biological and digital (as well as statistical and structural, to be discussed).

Theorists and researchers often look to the structure and organization of ancient biological systems for a pre-human understanding of CI (Clark, 2003; Fisher, 2009; Johnson, 2001), and the ant colony is the classic model. The ant colony can sustain itself for nearly fifteen years and includes thousands of individuals, and yet there is no leader or decision maker at the top of its social hierarchy. In fact, the concept of a social hierarchy is contrary to understanding the functioning of its CI. The colony is a bottom-up, not a top-down system where complex goals are achieved by the individual actors making simple decisions based on their own genetic programming and limited world experience. There is no macro awareness by the individual of the system's health as a whole, though by each individual following its own small role, the collective grows smarter over time and can respond to the changing needs of the environment. This is the process of *emergence* as explained by Johnson (2001), a process where "relatively stupid elements … think locally and act locally, but their collective action produces global behavior" (p. 18, 74).

The emergence of systems intelligence in biological settings like the ant colony is similar to the acumen seen in online digital systems. Clark (2003) cites the chemical trails of slugs and ants in his discussion of online consumer website design. Slug "goo" and ant pheromone trails are not just organic afterthoughts, but rather multifunctional system coordinates that work through the mechanism of *stigmergy*: "Major trials cut collective locomotion, costs, convey potentially useful information to new travelers, and perform a kind of farming function to boot" (Clark, 2003, p. 144). Such models of collective information and communication are adapted for digital life by such retailers as Amazon.com and search engines like Google; Clark (2003) labels this “swarm intelligence,” but the principles are the same as that of Johnson’s emerging systems—simple actors creating complex systems (p. 146). In this case, “vendors need never know” how the trails are laid (Clark, 2003, p. 147); again, macro-level awareness is not important. The major strength of this model, decentralization, may also be a major weakness, however, as Galloway (2004), Lessig (2006) and others have shown how decentralized and distributed organizational structures still fall prey to a dominant few through the implementation of protocols.

Statistical and behavior models of CI are equally as attractive and compelling as the biological models when considering digital media, and these models offer insight into why CI supersedes individual intelligence on wiki-based platforms. Sunstein (2006) discusses the major tenets of these models, specifically citing the Condorcet Jury Theorem and prediction markets as useful for the online context. The Condorcet Jury Theorem is a political science theory to explain group decision making. Sunstein (2006) explains:

> Suppose that people are answering the same question with two possible answers, one false and one true. Assume, too, that the probability that each voter will answer correctly exceeds 50 percent. The Jury Theorem says that the probability of a correct answer by a majority of the group increases toward 100 percent as the size of the group increases. (p. 25)

The converse holds true as well; if the majority of people have less than a 50 percent probability of answering correctly, the group probability moves toward zero. The implication of this postulate is that a collective becomes more intelligent—able to answer a question correctly in this case, or solve a problem in a broader context—when the right people are added to the group. The Condorcet Jury Theorem can be proven mathematically and is a steady instrument for measuring probabilities. Its limitations, however, arise when applied to human situations. Human behavior and decision making are often influenced by bias, self-consciousness, and the need to conform. Even in a well-informed group of individuals, these complications can render the theorem irrelevant.

Sunstein (2006) then proposes prediction markets as a more realistically useful model for an organization to gain access to "many minds" (p. 104). A prediction market works much like any market where values fluctuate based on actors' confidence in commodities, except here the commodity is information. A number of corporations, universities, and government departments have implemented prediction markets in their decision-making structures, and they:

> give people the right incentive to disclose the information they hold. … Recall that under the Condorcet Jury Theorem, the average vote of a large group will be wrong if most group members are likely to err. In a prediction market, the existence of incentives greatly increases the likelihood that each investor will prove to be right. Those without information will not participate; those with a lot of information will participate a great deal. (Sunstein, 2006, p. 104-106)

Thus, prediction markets control for the major weaknesses of the Jury Theorem, but limitations still exist in these statistical models. Although they do reflect collectivities and idiosyncrasies of human behavior and can predict facts with extreme accuracy, markets are not intended to handle issues of morality or judgment of ethical value: "For many of the most important questions that societies face, prediction markets will not be adequate, even if they incorporate the views of many minds and hence produce a ton of information" (Sunstein, 2006, p. 143).

Pierre Levy's (1997; 2001) notion of CI and its use for digital media and technology serves to infuse a humanism into the concept. Without fully discounting the aforementioned models, Levy (2001) cheerleads a CI that steps beyond the other models and proves more useful in human enterprises:

> The extension of cyberspace will transform the constraints that have dictated the range of possible solutions to political philosophy, the management sciences, and our organizational traditions in general. A number of these constraints have disappeared with the availability of new tools of communication and coordination. We can now envisage radically new ways of organizing human groups and relationships between individuals and collectives, which lack both historical models and precursors in animal societies. (p. 112)

The project of discerning and spreading knowledge is one dependent on human relationships, and in an early treatise on the Internet, Levy (1997) proposed fixing this social bond at the center of digital media, especially cyberspace. In theoretical response to the weaknesses of biological and statistical models of CI, Levy holds up the "ethical and aesthetic dimensions" of the concept as equally important as the "technical and organizational aspects" (p. 10). His approach to this work mirrors his interest in the humanistic angle, as he draws on parables from the Bible, lessons from art, and cases of moral judgment; he offers chapters on the ethics, the economy, and the atheology of CI.

For Levy, establishing CI is a wholly political project. In critiquing digital engineers' interest in biological models, he asserts that *his* conception of CI "must not be confused with totalitarian projects involving the subordination of individuals to transcendent and fetishistic communities" (Levy, 1997, p. 16). While appreciating the organization of the ant colony and the "emergent behavior that is globally intelligent" in that structure, Levy (1997) dubs the colony "prehumen," as it does not lead to the meaningful knowledge spaces he hopes new technology will move toward (p. 16). He also addresses statistical and market models, highlighting the parallel between their emergence in the 17th century and the development of Western capitalist expansion.

“Qualities are reduced to quantities,” he writes, going on to claim that while statistical science has “obvious utility” for governments and economies, “they provide us with a poor understanding of the nature of singularity as such, of events and their effects, dynamic configurations, anything, in fact, that is part of the world of signification in general” (Levy, 1997, p. 190, 203). A champion of the virtual (defined as “the possible,” not “the unreal”), Levy clearly worries that an intense interest in the commodity space of digital platforms like the Internet will reduce interest in the possibilities for its knowledge space.

Also noteworthy here is the concern for the individual in a system, something Johnson (2001) and the emergence theorists stress, but in a much different light. Johnson prizes the simple individuality of the actor, while Levy (1997) seems protective rather of the potential and freedom of the single actor. In this fundamental difference is the tension of competing ideas over which model(s) should inform digital projects of CI. Whereas Johnson steps back, claiming that “understanding emergence has always been about giving up control, letting the system govern itself as much as possible, letting it learn from the footprints” (p. 234), Levy steps forward, indicating that “this project implies a new humanism that incorporates and enlarges the scope of self knowledge into a form of group knowledge and collective thought” (p. 17). This debate between governance and freedom is one playing out on many new media platforms, including Wikipedia.

As the previous literature indicates, there are numerous, varying conceptions of what intelligence is and how it manifests itself in humans, technology, and collectives. Interestingly, lines of thought on CI posit that individual intelligence is largely unnecessary for certain systems to display intelligent organization and problem solving abilities, begging the question for this project: Do individual contributors, both human and bot, display intelligence on Wikipedia?

Intelligence on Wikipedia – Human and Otherwise

One of Wikipedia's core content policies that guides acceptable contributions to the site is "No Original Research":

> The term 'original research' is used on Wikipedia to refer to material—such as facts, allegations, and ideas—for which no reliable, published source exists. This includes any analysis or synthesis of published material that serves to advance a position not advanced by the source. (Wikipedia, 2012bb)

In concert with the other core content policies, "Neutral Point of View" and "Verifiability," the dictate for no original research is meant to ensure that Wikipedia maintains a level of objectivity to bolster its credibility as a reference source. Unlike scientific journals, literary magazines, or even the mass media, all of which attract their readers by featuring new and unique ideas, Wikipedia trades in established intelligence … at least in theory.

In practice, though, Wikipedia is the site of continuous new (and old) intelligent activity. Recall Gottfredson's (1997) broad definition of intelligence: it involves "the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience" (p. 13). Thousands of human contributors do each of these on the site each day, from Talk page discussions and debates around content, to deciding if edge cases sufficiently justify policy adjustments, to bot owners defending the function and utility of their bots, to new users figuring out how to edit their first page. The list could go on, as Wikipedia is in many respects a platform for problem solving, both at the global level (how do we get to the sum of all human knowledge?) and the extremely local level (how do you and I agree on the representation of this information?).

As described in Chapter V, technology plays a major role in the intelligent activity on Wikipedia as well, co-constituting and continually conditioning a dynamic system of actors and forces that maintain and develop the site. In the face of advancing artificial intelligence, chatterbots and spambots, Web 2.0 customization, targeted digital advertising, and the googlization of everything (Vaidhyanathan, 2011), it is tempting to consider the technologies on Wikipedia intelligent, specifically bots. Such a view, though, has sparked sentiments of technophobia among the site's general community. User and bot operator Tedder commented:

> I think people are scared of bots, perhaps a little too much. … Sometimes the things being discussed wouldn't be an issue except the word 'bot' is attached to it. I mean, editors make lots of edits without oversight, but bots making the same edits get a lot of oversight.

Many other bot operators expressed similar frustrations at the resistance bots often face in the larger Wikipedia community, and no time was more contentious than the public emergence of adminbots around 2007. An adminbot has the same access and privileges of an administrator, including the ability to delete pages and block users. To that point in the project's history, adminbots were against policy, as deleting and blocking were seen as actions too sensitive to be performed without human review. Adminbots were indeed on the site, though, being run from certain administrators' accounts. User Cyde called adminbots "an open secret," and User Snowolf referred to the log of administrator action statistics, some with hundreds of thousands of deletions, remarking: "As you see by the numbers, there is/was plenty of automation going on." Cyde recalls:

> People were un-ironically citing Terminator as an argument against adminbots. … Some of the people arguing against this stuff simply had no idea how programming works. I guess they thought it was like it was in the movies. But the bot isn't going to do anything that you don't explicitly program it to do.

Ultimately, a policy was developed to allow select adminbots who fit certain criteria, and currently nine bots roam the English WP with these privileges (Wikipedia, 2012m). But the conflict around adminbots indicated a fundamental misconception from the broader Wikipedia community about the intelligence of bots.

Held up to Gottfredson's definition, Wikipedia bots are hardly intelligent actors. Although User Josh Parris commented that "some bots do an amazing job, exhibiting something almost resembling judgment," he and virtually all of the interviewees see their creations as human-made problem-solving tools, only as smart as their code.[62] Advances in artificial intelligence are completely outside their domain, and barring the stray bot comment that fools a newbie user of the site, no Wikipedia bot would pass a Turing Test. User Multichill, a bot operator on many Wikimedia projects, sums up this dominant position from within the bot community: "I've never seen an intelligent computer, and I don't think I'll ever meet one. It does contain a lot of information, but it's not real intelligence. The wisdom is in the people using it."

What we find on Wikipedia, then, are masses of usually intelligent human users and preprogrammed, fundamentally unintelligent bots interacting on a regular basis; the implications of this heterogeneous population for CI will be discussed later in the chapter. First, though, we look at how the conceptual frame of CI is understood by those in and around the project.

---

[62] Josh Parris specifically referred to ClueBot NG, an anti-vandalism bot unlike most others. Rather than using simple heuristics and blacklisted words to detect potential vandalism, ClueBot NG uses "a combination of different detection methods which use machine learning at their core," including Bayesian classifiers, an artificial neural network, threshold calculation, and post-processing filters (Wikipedia, 2012j). Alluding to its sophistication, Josh Parris noted: "Recently I saw it picked up an edit about someone's 'meat flute.'"

The Frame of Collective Intelligence for Wikipedia

As previously discussed, the concept of collective intelligence can be applied to many different kinds of systems and actors, both material and immaterial, and has been used to understand behavior for over a century. Still, much of CI's recent applications have been in the context of digital media, and as Hopper (2008) points out, "The widespread proliferation of online participatory systems such as wikis and blog networks helped popularize the idea of collective intelligence" (p. 245). Wikipedia is a common example used in this discourse, and CI has been applied as a frame to understand collaboration and cooperation on the site. This section will explore how the CI frame has been used (or not used) by three interrelated discourse communities: the mass media, new media researchers, and the Wikipedia community itself.

In the Media

The tenth anniversary of Wikipedia in January 2011 was met with a bevy of media coverage, often featuring interviews with cofounder Jimmy Wales and reflections on major strengths and weaknesses of the site. On reasons for the site's success, Ash (2011) writes that Wikipedians offer several explanations:

> [The site] arrived fairly early, when there were not countless sites for fledgling netizens to spend time on; an encyclopedia deals (mainly) with verifiable facts rather than mere opinions, the common currency and curse of the blogosphere; above all, Wikipedia struck lucky with its communities of contributor-editors.

Indeed, the community of editors is commonly discussed in these stories as the principal catalyst for Wikipedia's success, with the openness and inclusiveness of the community's ethos outshining attention to those same qualities of the site's technology. However, discussions of how exactly a community of amateurs spread across the globe and primarily communicating anonymously online—the mechanisms of collaboration that cause the project to emerge as a highly functioning system—are almost completely neglected. Across ten major media outlets[63] covering Wikipedia's anniversary, only *The Atlantic* offers insight into this collaboration using the frame of collective intelligence, and it does this by offering short essays and editorials by prominent communication scholars and practitioners like Jay Rosen, Clay Shirky, and Yochai Benkler.

Though far from a scientific sample, this news coverage offers a snapshot of the broader media discourse around Wikipedia. Many of the concepts and buzzwords that permeate the literature on new media—CI, crowdsourcing, emergence, wisdom of the crowd—do not find their way into mainstream explanations of the site. Instead, this coverage focuses on major milestones (for example, growth in the number of articles in the encyclopedia), major controversies (incorrect information, or public figures caught editing their own profiles), and the increasing resonance of the site in popular culture. Descriptions of the technology behind the site are largely lacking, nay the lone line about how a wiki allows anyone to participate. Bots and automated tools are totally unrepresented. The average news consumer is left with the impression that Wikipedia works because it has a large community of contributors, each who offer a little portion of intelligence; beyond this, the workings of the site are left uninterrogated.

In New Media Research and Theory

Far different from the mass media, new media theorists and researchers often apply the concept of CI to Wikipedia, usually within the framework of explaining how the technical structures of the Internet enable collaborative behavior that leads to emergence. This is where a fundamental understanding of packet switching and the

[63] National Public Radio, *BusinessWeek*, Canadian Broadcasting Corporation, *The Guardian* (U.K.), *The New York Times*, *The Sydney Morning Herald* (Australia), *The Atlantic*, British Broadcasting Corporation, *The Daily Telegraph* (U.K.), *The Chronicle of Higher Education*

network traits of the Internet becomes important, as these elements become fully ingrained in the social theory put forth by many of these scholars. As both Baran and Davies emphasized in their models, packet switching produces an end-to-end distributed network, where power (or intelligence) is not concentrated in any one centralized node; instead, each node has a small amount of intelligence (relative to the entire system) and the ability to contribute those resources to the system via a number of paths. Lessig (2006) adds, "The end-to-end principle is a design philosophy about how networks should be built. It counsels that a network should be kept as simple as possible and that the intelligence required in a network be vested in the edge, or ends of the network, at least so far as possible" (p. 111). Terranova (2004) calls this "fringe intelligence," which "produces a space that is not just a 'space of passage' for information, but an informational machine itself—an active and turbulent space" (p. 67). Fuchs (2008) argues that from this technology-enabled space, self-organizing systems emerge, and he cites Wikipedia as "a dynamic, permanently changing communication system that grasps the characteristics of the Internet and online communication" (p. 318). Further, Terranova (2004) and Fuchs (2008) each emphasize how the network dynamics of human and technological actors materialize Marx's "general intellect" both broadly in the form of the Internet and specifically in the form of platforms like Wikipedia.

Others reinforce these connections between the end-to-end structure of the Internet, collective intelligence, and Wikipedia. Zittrain (2008) argues:

> Like the development of the Internet's architecture … Wikipedia's original design was simultaneously ambitious in scope but modest in execution, devoted to making something work without worrying about every problem that could come up if its extraordinary flexibility were abused. It embodied principles of trust-your-neighbor and procrastination, as well as Postel's Law, a rule of thumb written by one of the Internet's founders to describe a philosophy of Internet

> protocol development: "Be conservative in what you do; be liberal in what you accept from others. (p. 134)

Benkler (2006) calls this "technical agnosticism," which leads to "social agnosticism," or the exploration of different forms on online communication knowing that the full system will be unknowable; this position on CI is akin to Levy's (1997) more cultural take on the concept. Bruns (2008) emphasizes the community over the technology in his take on CI and Wikipedia, writing: "collective intelligence in its full form only emerges where at least a semi-organized collective has been established" (p. 111). A related but distinct strain of literature on collective intelligence and online communication has looked at the possibilities of emergent intelligence and cultural production for business, the economy, and the entertainment industries (Jenkins, 2006; Surowiecki, 2004; Tapscott & Williams, 2006), though from a less critical perspective than the aforementioned scholars.

A number of prominent critics have voiced skepticism for the emphasis that many have put on the idea of collective intelligence, including when applied to Wikipedia. Perhaps easiest to knock are the business manifestos touting the power of user-generated content and the democratizing element of participatory media. Van Dijck and Nieborg (2009) point out how such Web 2.0 cheerleading erases "the distinction between collective (non-market, public) and commercial (market, private) modes of production [and] between producers and consumers," as well as replaces the discourse of profit-oriented industrial production with the more positive and empowering discourse of peer production (p. 856). Lanier (2006; 2010) extends this critique to most of the rhetoric around social media and online knowledge production, arguing:

> The problem is in the way the Wikipedia has come to be regarded and used; how it's been elevated to such importance so quickly. And that is part of the larger pattern of the appeal of a new online collectivism that is nothing less than a

> resurgence of the idea that the collective is all-wise, that it is desirable to have influence concentrated in a bottleneck that can channel the collective with the most verity and force. This is different from representative democracy, or meritocracy. This idea has had dreadful consequences when thrust upon us from the extreme Right or the extreme Left in various historical periods. The fact that it's now being re-introduced today by prominent technologists and futurists, people who in many cases I know and like, doesn't make it any less dangerous.

Lanier (2006) is particularly weary of CI's reliance on technology—algorithms and programs—to coproduce knowledge, concluding: "History has shown us again and again that a hive mind is a cruel idiot when it runs on autopilot." Finally, another prominent strain of critique around CI on Wikipedia is that the knowledge produced regresses to the mean, producing mediocrity, rather than rises to the top (Carr, 2005; Keen, 2008).

In the Wikipedia Community

How do people from within the Wikipedia community understand CI, and do they consider the project to be an example of the concept? The WMF stays away from the rhetoric of "collective intelligence," "wisdom of the crowd," "crowdsourcing," and related terms that may be interpreted as buzzwords. Instead, they prefer to use language focused around "knowledge" and "collaboration": for example, the "Letter from the Directors" in the WMF's most recent annual report describes Wikipedia as "the most important collaboratively created repository of knowledge in history" and "the most powerful example of volunteer collaboration and open content sharing in the world today" (Wikimedia Foundation, 2011a). However, the idea of CI comes through in their official communications in other ways; that same annual report includes the following quote from an Indian advisory board member:

> Wikipedia is the only place that's allowed for a system where generosities can be coupled and multiplied and leapfrogged upon, where therefore one's individual generosity, the fruits of it and the results of it are something that are just far beyond the effect of that act alone. (Wikimedia Foundation, 2011a)

The idea that collaboration on Wikipedia builds upon itself to create a structure that is broader and deeper than the sum of its parts does not appear to be resisted by the WMF, though it is not proclaimed either.

The WMF is fond of promoting their encyclopedia as an effort to aggregate "the sum of all human knowledge," prominently featuring the phrase in an aspirational context in many of their publications: "Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge." By promoting collective knowledge, not collective intelligence, the organization is playing it safe in a sense, sticking to the more factual and more tangible aspects of mass collaboration. But collective knowledge and collective intelligence are not technically equivalent, as the latter is a much more dynamic concept, suggesting an ability to solve problems and achieve complex goals through small, localized actions and decisions. By staying away from the CI discourse, the WMF is actually distancing itself from what many of the aforementioned scholars find so powerful about wiki-based technology: through distributed action, a system can outperform others based on centralized control.

At a more individualized level, Wikipedia contributors are divided in their assessment of collective intelligence on the site. Bruns (2008) found that administrators are "surprisingly dismissive" of the CI tag for Wikipedia, a sentiment partly found in the present research. A slight majority of interviewees said they believe Wikipedia is an example of CI, including some who strongly identify with the concept. A number of others, however, were unsure the tag quite fits, instead preferring to call the project "collective knowledge," "collective writing," "crowdsourcing," or even "a (mostly) functional system of mob rule." A handful of interviewees felt CI is a buzzword that somehow inflates what actually happens on Wikipedia, elevating the intangibleness of the project above the actual work editors contribute. User Josh Parris equates Wikipedia with a brick-and-mortar analog: “Some editors do anti-vandal work, others write new articles, and others do maintenance. Organizations often have security guards, workers and janitors. If there weren’t computers involved, would a few thousand people editing an encyclopedia be regarded as collective intelligence?” Indeed, in this comment we see the critiques of Wikipedia’s claim to CI, but we also see the importance of digital media and ICTs to our understanding of the concept. Josh Parris is right; online technology has fundamentally changed how we think about work, collaboration, problem solving, and communication.

Collective Intelligence on Wikipedia

To truly explore whether Wikipedia is a system of collective intelligence, we need to consider the specific mechanisms and models that suggest such a classification. Again, many of these concepts originate in the natural sciences, where they have been observed in systems of social animals, but researchers teasing out the possibilities of CI in human and technical systems are also turning to these ideas to understand particular phenomenon (Salminen, 2012; Tovey, 2008). Three of the most foundational concepts—stigmergy, distributed cognition, and emergence—are examined here in the context of Wikipedia.

Is there Stigmergy on Wikipedia?

A key concept for understanding the mechanism that makes collective intelligence possible is stigmergy. Originally proposed by French entomologist Pierre-Paul Grasse in the 1950s, stigmergy served to explain why a termite colony could develop into a complex system despite the relative simplicity of each individual termite; the insight here was that coordination emerges from each termite leaving a trace (a pheromone) in the environment when it successfully completed a task (moving a bit of mud in the colony) (Bonabeau, 1999; Gregorio, 2003). Others would be attracted to the more successful areas, and in turn, leave their own trace, reinforcing that preference. The pattern would continue until another individual finds an alternative success, at which time the pattern could shift and the old traces could eventually fade away. Grasse (1959) wrote:

> The coordination of tasks and the regulation of constructions does not depend directly on the workers, but on the constructions themselves. The worker does not direct his work, but is guided by it. It is to this special form of stimulation that we give the name STIGMERGY (*stigma*, mark; *ergon*, work).

Overall, this creates a system of indirect coordination where two individuals likely never interact directly, and yet collectively carry out complex tasks. The mechanism of stigmergy is used in entomology to explain the construction of large, dynamic ant hills and the establishment of shortest-path ant trails to a food source.

Outside of biology, the idea of stigmergy, or indirect coordination in large-scale systems, has been applied in the fields of artificial intelligence, robotics, and Internet studies (Bonabeau, 1999; Clark, 2003; Elliott, 2006; Johnson, 2001). Of the latter, Gregorio (2003) writes, “The World-Wide Web is the first stimeric [sic] communication medium for humans,” as no other medium has created an environment where everyone is able to read and write messages in a public way that is available to everyone else. The blogosphere, Google’s PageRank system, eBay, Slashdot, and other online recommendation systems are some examples of websites that operate with the mechanism of stigmergy; masses of individuals are communicating indirectly, and yet they are learning from each other, making decisions based upon others’ traces, and enabling complex systems to emerge.

Elliott (2006) has looked specifically at the wiki to investigate online stigmergic collaboration. Recognizing that any form of collaboration is both social and dependent on communication ("collaboration cannot be a solo venture"), the author posits that small group collaboration of less than 25 actors relies heavily on social negotiation through direct communication (Elliot, 2006). Amongst larger groups, however, direct communication is impractical and unfeasible; collaboration at a mass level, then, must rely on indirect communication, and the wiki facilitates this by allowing unrelated actors to leave their trace (edits, comments, etc.) in the environment. Elliot (2006) suggests this is stigmergy in action, with strong benefits for the overall production of the system:

> The use of stigmergic communication to sidestep social negotiation effectively fast-tracks the creative gestation period, removes social boundaries and as a consequence lowers the 'costs' of contribution by eliminating the need to become acquainted with and maintain relationships with fellow contributors.

This is the main difference between collaborative authorship, the term used by some interviewees to describe Wikipedia, and stigmergic collaboration: the emphasis on personal relationships and direct communication essential to the former is unnecessary (though still possible, in some forms) for the latter.

When we look specifically at the sociotechnical elements of Wikipedia, we find a system constructed to benefit from stigmergy. Technically, the system allows contributions from virtually anyone, be they registered users, anonymous users, power users, or one-time, drive-by users. Every action leaves a trace (an edit), and every trace is publicly documented and displayed for all to see. What's more, traces can be reinforced by the use of hyperlinks, which create trails to other traces. These technical qualities support and enable semantic mechanisms of stigmergy. Article content and Talk pages can be edited by anyone, and the traces of previous edits elicit new traces by other actors, who can reinforce or ignore the traces that came before. This is "variation and selection at work: different people contribute different text fragments, some of which are clear, accurate and relevant, some of which are less so," (Heylighen, 2008, p. 309). What this enables is the development of meaningful content, policies, and discussions by a large number of diverse, unrelated, and geographically dispersed actors who sometimes have no direct communication with each other.

We know, however, that many Wikipedia contributors do communicate directly to discuss and debate content and in many ways get to know each other at a social level. These personal relationships do not contradict the stigmergy thesis, but rather work with the mechanisms of stigmergy to further strengthen the overall coordination of the site. Remember, stigmergy operates on a macro level, accounting for the indirect communication of diverse editors, whereas sub-communities of editors who know each other and form a coauthorship model operate on a micro level. Elliott (2006) suggests that "stigmergic collaborations naturally form clusters representing the contributors' interests [… and] these 'contributor groups' form networks that may operate either implicitly or explicitly, with groups actively working together or remaining largely unknown to each other." Thus, we can understand collective intelligence on Wikipedia as the result of a stigmergic sociotechnical system that embraces pockets of social cohesion and negotiation.

Is there Distributed Cognition on Wikipedia?

Another concept proposed for understanding collective intelligence on Wikipedia is that of distributed cognition (Geiger & Ribes, 2010; Jesus, 2010). Largely originating in the work of Hutchins (1995) on how a host of people, artifacts, and tools are necessary in keeping a U.S. Navy ship on course, the framework of distributed cognition suggests that human knowledge and cognition is not confined to the individual mind, but it is distributed amongst people and objects in one's environment. As such, cognitive processes like memory, problem solving, and decision making are largely dependent on systems of actors and representations. Distributed cognition shares a number of ontological values with actor-network theory, namely that knowledge and understanding are the products of networks composed of both human and non-human actors. The theory has been taken up by many in the fields of human-computer interaction and computer-supported collaborative work, as it has "a special role to play in understanding interactions between people and technologies, for its focus has always been on whole environments: what we really do in them and how we coordinate our activity in them" (Hollan, Hutchins, & Kirsh, 2000, p. 174).

Jesus (2010) explores the framework of distributed cognition on Wikipedia, claiming it can be useful for thinking about how Wikipedia is written, and "how to account for the many tinkering edits and the fewer substantial additions of content" (p. 1). The researcher finds that the distributed cognition of editors and articles can be broken down further into the cognition for planning (reflective, thoughtful decision making) and the cognition for improvising (reflexive, instinctual decision making). Overall, Jesus (2010) claims that "cognition is emergent in these socio-technical systems [wikis]" (p. 208). Geiger and Ribes (2010) look specifically at the way cognition, or the ability to recognize and ban a vandal on Wikipedia, is distributed across a number of editors, bots, and semi-automated tools. By making certain evaluative processes "mundane," semi-automated tools like Huggle and Twinkle boost the cognitive capacity of human editors, thus enabling “a form of distributed cognition among otherwise disconnected vandal fighters” (p. 8). Furthermore, the authors argue, “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” (p. 7). Overall, the distributed cognition of these heterogeneous actors maintains a social order on Wikipedia that is largely attributed solely to its human bureaucracy.

Other examples revealed in the present research offer additional windows for understanding distributed cognition on the site. It would be quite inaccurate to attribute the notions of artificial intelligence explored earlier in this chapter to Wikipedia bots (though some may indeed pass the Turing Test for new and inexperienced contributors). To understand distributed cognition, though, we need not attribute powerful intelligence to these software tools; we merely need to recognize how, even in their simple state, they bear some of the cognitive load of problem solving and decision making in the network of actors. DeltaQuadBot, mentioned in Chapter V, is a prime example. DeltaQuadBot warns administrators of possibly inappropriate usernames by reporting them, along with an explanation, to an administrator noticeboard. The bot both recognizes these usernames and offers a context-specific explanation based on a blacklist and whitelist of words generated by its previous experiences; this is a basic keyword match, achieved with a mere handful of lines of programming code. For example, it may return a match for the string “homo” or “bureaucrat” with the following notes (respectively):

> LOW_CONFIDENCE,NOTE(Homo is Latin for "man" and Greek for "same". Please keep this in mind when evaluating a name.)
>
> NOTE(Name may imply a position of authority),WAIT_TILL_EDIT

But by completing this task and offering information that will aid the administrator who decides on a potential block, the bot is significantly contributing to the decision process for the case, relieving the admin of certain cognitive responsibilities (i.e. knowing the possible cultural or linguistic allusions in a name), and ultimately shaping the social, discursive space of Wikipedia. As Geiger and Ribes (2010) summarize, “Semi- and full-automated tools constitute an information infrastructure that makes possible the quick and seamless processes of valuation, negotiation, and administration between countless editors and issues” (p. 2).

Is there Emergence on Wikipedia?

The concept of emergence is closely linked with stigmergy and distributed cognition, and each helps us flesh out the larger, sometimes ideologically-charged idea of collective intelligence. Whereas stigmergy is a mechanism that produces coordination, and distributed cognition is a way to understand the coordinated problem solving between actors and their environment, emergence can be conceptualized at a more macro level. The essence of emergence lies in the proverbial notion “the whole is greater than the sum of its parts”; more properly, it can be defined as “a global property [that] cannot be determined from knowledge of its components” (Johnson, 2008, p. 271). Christakis and Fowler (2009) use a familiar analogy to explain the concept:

> A cake has a taste not found in any one of its ingredients. Nor is its taste simply the average of the ingredients’ flavors—something, say, halfway between flour and eggs. It is much more than that. The taste of a cake transcends the simple sum of its ingredients. (p. 26)

Physicist Doyne Farmer said of emergence: “It’s not magic … but it feels like magic” (as cited in Corning, 2002, p. 18).

Indeed, emergence has been used to explain complexity in a number of fields, including physics, psychology, philosophy, economics, art, artificial intelligence, and computer science. The now classic example for demonstrating emergent behavior is that of cellular automata. Using a mathematical grid, basic rules of behavior are given to each cell, with rules relative to the state of neighboring cells (for example, turn "on" if two neighbors are turned "on"). These rules are defined at the starting state and remain constant once the system is put in motion (Figure 40). Amazingly, such simple rules and binary actors, when run through numerous iterations, begin to form complex patterns of motions and groupings that resemble certain biological behaviors (Terranova, 2004). Today, computer simulations explore cellular automata in the discipline of artificial life, which strives to understand emergence in complex biochemical systems by imitating them digitally.

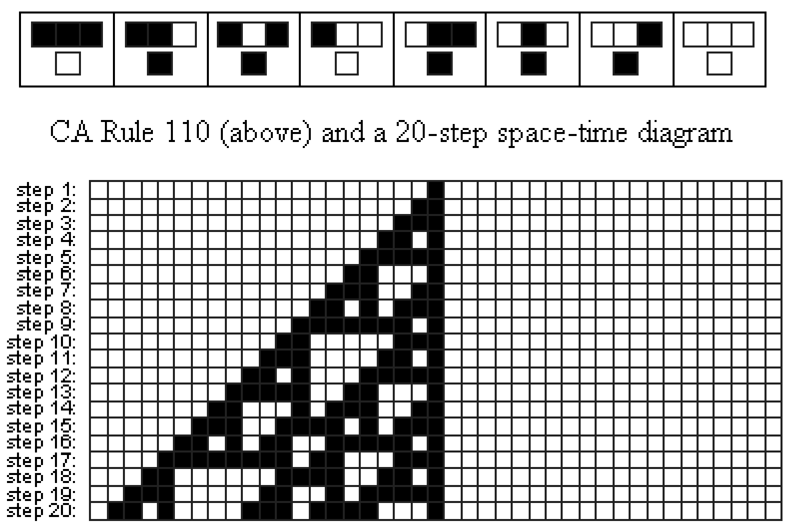

Figure 40. Example of cellular automata rules, steps, and patterns. Image from H. Sedaghat.

Fuchs (2008) argues that emergence is a fundamental quality of self-organizing systems, including social systems ranging from ants to humans. Of the latter, Johnson (2008) highlights the fall of the Berlin Wall as a prime example:

> It was not predicted, nor was it planned in any localized sense: the individuals that participated in the process that led up to the event never had that goal, nor knew that this was a possible outcome of their activities. It just happened as an emergent CI solution to a collective problem. (p. 272)

The author further suggests that history is marked by many similar events, though they are often attributed to individual efforts, as "historians are not generally appreciative of CI" (Johnson, 2008, p. 272). Investigating social behavior from this perspective, Christakis and Fowler (2009) suggest that "social networks have emergent properties …. A social network is a kind of human superorganism, with an anatomy and physiology—a structure and a function—of its own. Our local contributions to the human social network have global consequences" (p. 26, 289). Online social behavior and networking, then, offer a unique sociotechnical environment to explore emergence; the era of Big Data and Post-demographics promises insights into how larger structures of experience emerge from local online actions. Recent research into digital folksonomies[64] has already revealed emergent properties (Robu, Halpin, & Shepherd, 2009).

So should we consider Wikipedia a site of emergence? It is certainly romantic to think so. With over 21 million articles across 285 languages, the project's mere size points to a large and diverse community of collaborators. The encyclopedia is a work of text likely impossible for one author, or one hundred authors, or even perhaps one thousand authors. But does this collaboration lead to a global property that is actually

[64] A portmanteau of "folk" and "taxonomy," a folksonomy is "a system of classification derived from the practice and method of collaboratively creating and managing tags to annotate and categorize content" (Wikipedia, 2012c). The term is synonymous with collaborative tagging and social indexing.

more than the sum of its individual parts? Functionally, we are likely to say that Wikipedia in its entirety is not emergent; in fact, the encyclopedia is not *more* than the sum of its parts, but *is* the sum of its parts. As Figure 21 from Chapter V depicted, Wikipedia is functionally a collection of articles or a sum of individual pages. If we consider an individual article itself, again we may fall into the trap of imagining its magical genesis from the minds of many contributors. In reality, the article is a precise sum of the additions and subtractions made by a finite number of contributors, and the fact that every edit and every contributor is documented to the hour and minute further demystifies its creation.

What does seem emergent in particular articles with numerous contributors is the voice of the writing. Voice is defined as the unique writing style and word choice that represents an author's attitude and character; it is often associated with particular authors, though voice can also be established by collectives (i.e. the voice of the organization) (Hacker & Renshaw, 1985). Since Wikipedia contributors have the ability to edit and rewrite existing material, the resulting content is an amalgamation of styles, often irreducible to a single author's voice.[65] Maxwell and Felczak (2008) describe the wiki as a "scaffolding technology" that produces a style that "subsumes individual voice," citing Wikipedia as a prime example.

As this dissertation project has argued, though, Wikipedia is a dynamic sociotechnical system, so we need to examine Wikipedia beyond its mere function and

---

[65] Some articles do maintain a strong element of the original author's voice, especially if the primary author becomes "WP:Owny" (Rich Farmbrough) of the article, a position discouraged by Wikipedia but a reality of the project nonetheless. The issue of article "ownership" is double-edged sword, according to User Ram-Man; these devoted contributors are often necessary for initially creating and building up articles, but they then become averse to other editors making changes, even if those changes are correcting something erroneous.

face value attributes in order to thoroughly investigate possible emergent properties. According to Fuchs (2008), online social networks like MySpace fit many of the characteristics of a self-organizing system, which itself implies the emergence of structure and order, and Goldspink (2010) hypothesizes that order on Wikipedia is indeed emergent. Wikipedia is a *complex* system of millions of human and technological actors, and no one actor—even its benevolent dictator Jimmy Wales—can fully comprehend or control the activity on the site. Information production is *dynamic* and *distributed* across the system; the states of content and social relations are constantly changing through localized interactions, and yet system-wide policies, procedures, and standards have emerged to produce *order* and *cohesion*. The consensus process itself, which guides nearly all collective decisions on the site, can be seen as emergent as well. Distinct from a voting system, where an ultimate decision is reducible to the individual votes, consensus decision making seeks to arrive at a common solidarity or agreement that may be negotiated from the individual opinions. Finally, Wikipedia is fundamentally *open*, meaning actors and content can come and go, constantly subjecting the system to *fluctuation*, *chance*, and *unpredictability*.

Still, it is difficult to fully conceptualize emergence on Wikipedia in the same way that philosophers and psychologists conceptualize consciousness as emergent from the brain and body. Chalmers (2002) labels this latter example *strong emergence*, and in fact cites consciousness as the only example where emergent high-level phenomenon cannot be deduced from lower-level truths even post facto. The author argues that all other examples of emergence are *weak emergence*, where the higher-level phenomenon is unexpected given lower-level truths, and yet can be understood as a result of interactions at the elemental level. Wikipedia seems to fit this latter categorization, as its aforementioned emergent characteristics, much like those of the cellular automata, can be traced to micro-level rules and activity. Still, even Chalmers (2002) admits this distinction may be semantic, as the concept is currently used in many different ways: "Typical uses of the term 'emergence' may well express cluster concepts with many different elements" (p. 12). As both Levy (1997) and Lessig (2006) argue in different ways, cyberspace is a fundamentally emergent reality, the product of both human and technological systems, and what is important about platforms like Wikipedia is that they provide the environment for new forms, social relations, and knowledge to emerge from cyberspace itself.

Collective Intelligence and Sociotechnical Systems

The previous section argued that there is sufficient evidence to label Wikipedia a system of CI based on established models for the phenomenon. These models are largely derived from fields outside communication and Internet studies, though this discussion pointed out unique ways that Wikipedia, as a digital media platform, both fits and advances conceptions of CI. In fact, the possibilities for CI opened up by ICTs are in many ways different than the biological analogs generally used to introduce concepts like stigmergy and emergence. Digital sociotechnical systems offer new complexities as well as opportunities for groups of actors to share knowledge, solve problems, and establish order. This section explores the implications of sociotechnical systems for the concept of CI and argues that Wikipedia exemplifies many of the aforementioned possibilities.

Communication is a fundamental element for unlocking these possibilities, and the wiki has established itself as a flexible, durable, and open sociotechnical object for managing communication among a large group of actors. Levy (1997), writing before the popular rise of wiki technology but seemingly anticipating its appearance, argues that large-scale CI "must be based on digital information technologies":

> New communication systems should provide members of a community with the means to coordinate their interactions within the same virtual universe of knowledge. This is not simply a matter of modeling the conventional physical environment, but of enabling members of delocalized communities to interact within a mobile landscape of signification. Events, decisions, actions, and individuals would be situated along dynamic maps of shared context and continuously transform the virtual universe in which they assume meaning. (p. 14-15)

Wikipedia provides such a system, a virtual universe of knowledge where a global community comes together to coordinate efforts and solve problems. Additionally, Wikipedia is dynamic, as interactions constantly maintain a tension between forces of change and forces of stasis that keeps the project whole and yet flexible, and the site is organically self-organized, with order predominantly emerging from internal energy. Wikipedia also offers a space for social communication; though much of the social interaction takes places within task-oriented projects (authoring articles, setting policies, recognizing good work, censuring inappropriate behavior), Wikipedia is an enormous space for social performance, from which a communal environment emerges. As Levy (1997) writes, "Collective intelligence is born with a culture and grows with it" (p. 16).

Sociotechnical systems like Wikipedia offer the ability to transcend systems of CI like the termite colony or cellular automata, which rely on what Christakis and Fowler call "zero-intelligence agents" (p. 25). As we've seen, the mechanisms of stigmergy can be set in motion by a population of simple actors with no global awareness. Emergent behavior can arise in a similar environment of agents that show no intelligence beyond following an algorithmic set of rules. Indeed, an irony of collective intelligence in natural and simulated systems is that it relies on little intelligence at the individual actor's level (relative to any human-level intelligence). But what if there was intelligence at the individual level?

The Internet is built as an end-to-end distributed system, pushing the responsibility for intelligence to the edges of the network, the individual users. This technical architecture creates a unique opportunity for CI beyond the basic models, as it allows for emergence that is facilitated by intelligence at the individual level. What we know as "the Internet" or "cyberspace" emerges from the communication of individual users across the network, but this CI is both structural and substantive, based on what Levy (1997) calls a "culturally informed intelligence" (p. 16). The author argues that a "new humanism" will be embodied in this CI, a cyberculture, much different than the social organization emergent from zero-intelligence agents:

> The intelligence of the group is no longer the mechanical result of blind or automatic activities, for it is individual thought that perpetuates, invents, and mobilizes that of society. … In place of the "invisible hands" of the termite colony, we have the visible hands and imaginable dynamic of expanding virtual universes. Through their interaction with diverse communities, the individuals who animate the knowledge spaces are, far from being interchangeable members of immutable castes, singular, multiple, nomadic individuals undergoing a process of permanent metamorphosis. (p. 17)

As this dissertation project has demonstrated, Wikipedia is built as a digital sociotechnical system that takes advantage of both the invisible and visible hands of self-organization. By relying on sociotechnical ensembles of human intelligence, programmed bots, social bureaucracy, and software protocols, the humanistic CI that Levy imagined is realized in a virtual knowledge space that embodies information as both product and process while empowering its community to explore the cultural possibilities of its collectivism.

Conclusion

From its original development for a small programming workgroup to its widest and most dynamic implementation on the world's largest encyclopedia, the wiki has served admirably as a tool for aggregating and organizing information and knowledge. This chapter has argued that the sociotechnical collaboration that takes place on a mass scale on Wikipedia produces something more: collective intelligence. Analysis and interpretation of activity on the site suggests the mechanisms for CI, the patterns and arrangements of actors and behaviors similar to those identified in other disciplines, are indeed present and produce an emergent cyberculture unique to the digital media environment. Though the moniker of CI is not always acknowledged by the Wikipedia community itself, this is only fitting, as the intelligence agents who maintain and grow the system are generally pragmatic in their work, developing generally unintelligent bots and tools to tackle some of the less appealing tasks on the site. However, the collective effort of these diverse human and technical actors, intelligent and otherwise, largely fits both our scientific and cultural definitions of CI and offers insight into the potentially revolutionary possibilities of ICT-enabled mass collaboration.

CHAPTER VII

CONCLUSION

In early 2007, Wikipedia broke into Alexa's Top-10 most popular global websites (Perez, 2007; Snow, 2007). In little over six years, a project that started as a lifeline for a failing, expert-written encyclopedia had grown into an online resource with nearly 10 million articles in over 250 languages. In April of that year alone, over 45 million unique Internet users would visit the site, reading about their favorite hobbies or celebrities, grabbing information or references for a research project, or perhaps settling a bet over who won the first FIFA World Cup back in 1930.[66] Wikipedia had arrived in both online culture and popular culture, with Stephen Colbert of *The Colbert Report* famously saying of the site, "It's the first place I go when I'm looking for knowledge, or when I want to create some" (Colbert & Hoskinson, 2007). With over five million registered editors and an untold number of anonymous contributors, the encyclopedia had largely weathered criticisms and controversies over inaccuracies (Helm, 2005), demonstrating the enormous potential of mass collaboration through digital media.

Today, Wikipedia sits comfortably at #6 in Alexa's rankings, with nearly 15% of all global Internet users visiting the site on a given day. But the project represents something much larger than its popularity as a place to go for information. In the turbulent Internet environment marked by debates over net neutrality and the encroaching influence of the media industries, Wikipedia is perhaps the best and certainly most visible example of what openness can achieve. The project stands alone among an array of privately held search engines and social networking platforms at the top of the Web's rankings, and unlike the abstract notion of the blogosphere, Wikipedia is a unified

[66] Host nation Uruguay defeated Argentina 4-2. Check out the article for more on the match.

entity—an endeavor with a mission, a community, an organization, and an end product: the sum of all human knowledge. Holding fast to many of the open-source ideals that continue to motivate its creators, while at the same time negotiating the largely commercialized landscape of the dot-com world, Wikipedia has become a legitimate online public good.

Using a theoretical framework informed by digital media studies, science and technology studies, and the political economy of communication, this study examined the material and ideological conditions in which this public good has developed and maintained itself. Specifically, this research explored the interplay between the technical and social actors and structures on Wikipedia, revealing a dynamic sociotechnical system of collaboration that creates a platform for the emergence of collective intelligence. This chapter discusses significant findings of the study, as well as its contributions to the field of communication studies and the broader, interdisciplinary field of Internet studies. The chapter concludes by discussing new questions and areas for further study indicated by this research.

Major Findings

Informed by previous research, this study posed four research questions to guide its exploration of Wikipedia's technical and social elements. These questions took a broad view of the project, but offered the opportunity for careful analysis of the history, details, and microstructures of the site. The following summary of findings is guided by these questions, which encapsulate the study's main theoretical concerns.

Research Question 1 asked: “What is the historical context for the technical development of Wikipedia? What actors and forces have shaped its code, community, and infrastructure?”

Wikipedia is the latest project in the longstanding human pursuit of a comprehensive reference work of the world’s knowledge, and like previous attempts, it is heavily conditioned by the social environment of its creation. Unlike most other attempts, however, Wikipedia’s social environment is itself a product of new technical innovations: digital media and the Internet. The mass user interaction that makes Wikipedia possible was envisioned by early pioneers like Tim Berners-Lee, developer of the World Wide Web, and Ward Cunningham, inventor of the wiki, but did not become a reality until online design, fostered by increasing access to and commercialization of the Internet, turned toward the centrality of participation. Wikipedia was launched in the wake of this online paradigm shift and has subsequently flourished because of it.

The project’s technical development has been shaped by the dueling ideologies of the open-source software movement and postindustrial capitalism. Wikipedia’s core software, MediaWiki, is open-source and largely written by volunteers ideologically aligned with the free and open-source software movement, which advocates for the collaborative production and free distribution of immaterial, intellectual works like programming code. Many of the extensions and bots that have been developed to work with MediaWiki are also open-source, and choices over the project’s history on how to scale the project with its increasing size and popularity have prioritized open-source options.

To successfully manage this growth and expand its material infrastructure, however, Wikipedia has become a strategic participant in the contemporary information technology economy. Organizing on a non-profit model, its parent Wikimedia Foundation has established relationships with technology firms, media corporations, and foundational benefactors to bolster its funding and expand its assets, namely the hardware and servers necessary to run the site. By capitalizing on its cultural resonance and unique position in the Internet ecosystem, Wikipedia has been able to strategically plan for the project's future while maintaining its mission, community of volunteers, and base of individual donors.

Research Question 2 asked: "In what ways is Wikipedia a sociotechnical system? What roles do the social and technical infrastructures of the site play in its collaboration and conflict? How does science and technology studies (STS) contribute to an understanding of the site's performance?"

The STS perspective is grounded in the belief that technology is thoroughly social, and to understand complex technological systems, we need to understand the activity, influence, and meanings given to all actors in that system. As such, STS is helpful for understanding the ecology of Wikipedia, one of heterogeneous actors working in concert and conflict to bring a dynamic system into a relatively stable state. Over time, both social bureaucracy and technical protocol have developed on the site, but neither has developed in isolation. From user rights and privileges being crystalized in MediaWiki's code to extensive policies on what sorts of automated tools are acceptable, Wikipedia is largely a sociotechnical system driven by the codependencies of its actors. Indeed, it is this sociotechnical nature, built up from the very concept of the wiki, that allows Wikipedia to maintain its obduracy in a continuous flow of activity.

Additionally, this study focused in on the roles of bots and bot operators in the sociotechnical system of Wikipedia. Bots, or automated software programs, are notable for their hybrid character, as they are in one respect technological entities, made of code and algorithmic in nature, but in another respect, they are social entities, enforcing bureaucracy and facilitating decision making. What's more, bots actually occupy a position in the layered structure of interaction with the site akin to human users, and bots are often personified and anthropomorphized in detailed ways. Still, most bot operators consider their creations mere tools, or unintelligent laborers whose work, while necessary to the health of the site, is fundamentally of lower value than that of human users. The STS perspective allows us to understand these incongruities by recognizing that both substance and meaning are constantly in flux in a digital media environment like Wikipedia, aligning and congealing in one way here and in another way there, but always the product of sociotechnical negotiation.

Research Question 3 asked: "In what ways is Wikipedia a system of collective intelligence? How is this frame applied and understood by the mass media? The research community? The Wikipedia community? How do other concepts of intelligence inform this discussion?"

Many media theorists and Internet researchers have been keen to apply the concept of collective intelligence (CI) to describe the collaboration that takes place on Wikipedia. They argue that the convergence of Internet principles like end-to-end intelligence with mass participation leads to an end product that is greater than the sum of its parts. The mass media generally passes over this conception of the project, instead focusing on its content and social formations, while the Wikipedia community itself is divided on the notion of CI, cautious of buzzwords and concepts that misrepresent the project or devalue the individual work that goes into it.

Still, this study finds that Wikipedia does fit some of the key models that suggest collective intelligence. Largely adapted from the biological sciences, concepts around CI are unique from traditional measures of human intelligence, which largely focus on the individual, and artificial intelligence, which largely aims to mimic human intelligence (though recent strains of research have moved away from this attempt). CI instead relies on a large number of actors, often without much individual intelligence, interacting with their environment in patterned ways.

The wiki itself is built for the mechanism of stigmergy, as individual contributors leave edits (or traces) of their work in the environment, which in turn directs those who come after them. Wikipedia's network of actors and content also enables a system of distributed cognition, where the responsibility of tasks is spread across a number of agents and structures. Finally, as a self-organizing digital system, Wikipedia creates the conditions for the emergence of complex superstructures irreducible to their individual parts. Together, these signs of stigmergy, distributed cognition, and emergence indicate that Wikipedia is a system of CI.

This CI is a bit different from the prototypical anthill of its biological applications, though. Wikipedia is a system of both intelligent (human) and unintelligent (bot) users, each necessary for the continued functioning of the project at its current scale. This is a heterogeneous, sociotechnical system of collaboration that produces a cyberculture, or culturally informed shared intelligence, unlike other systems of CI and unique to the digital media context.

Research Question 4 asked: “What can a social network perspective reveal about the collaboration between human and non-human actors on Wikipedia? How could this collaboration be visualized? What implications emerge from this picture?”

Though unlike other popular platforms like Facebook and Google+, where making connections to other users is the ultimate goal, Wikipedia can and should be considered a social network, as users collaborate around common interests, build relationships, and acquire status and reputation. The project’s core software is not structured around user-to-user connections the way the aforementioned services are, but the very nature of the wiki is to create links between content, including the communication between contributors. Groups of users develop around shared pursuits, with this study focusing in on a particular subcommunity, bots and bot operators, who standardize practices, organize activities, share experiences, and regulate behavior.

Moreover, Wikipedia should be considered a social network that encompasses both human and non-human users. A generally unconventional assertion, though one supported by the major tenets of actor-network theory, bots are significant actors in the network of collaboration, working with and communicating with their material and immaterial counterparts. As this study’s sociotechnical analysis revealed, human users and bots are intimately related in the network of Wikipedia, so by removing bots from the picture, we eliminate an important element for understanding power and influence in the system. Social network analysis helps us both conceptualize and visualize the relationships between all users in the creation of Wikipedia content, and the exploratory analysis in Chapter V revealed a remarkably similar centrality between highly active human and bot contributors. However, the work carried out by these various users falls along a spectrum of interpretation, indicating that the social network perspective should not be applied in isolation from other empirical and theoretical positions.

Contributions of the Study

This study makes a number of valuable contributions to the field of new/digital media studies, including the subfields of computer-supported collaborative work and wiki studies. Theoretically, this research approached Wikipedia from a perspective informed by both science and technology studies and the political economy of communication. Mosco (2009) and Fuchs (2011a) have called for more critical work that combine these traditions, as both employ a realist epistemology interested in the relationship between materiality, knowledge, and social practice. Much of the existing research on Wikipedia, however, has featured either strictly positivist or social constructionist orientations, focusing on the site's technology or its user community. By examining the project with a deep attention to issues of power and control, this study was more thoroughly able to investigate how the social, material, and immaterial worlds of Wikipedia hold each other in place. This allowed the concept of a sociotechnical system, which has traditionally been used in a functionalist way, to be extended to a critical analysis of how digital media can subsume conventional and established notions of the social and the technical.

In addition, this study contributes to the theoretical understanding of collective intelligence in online media systems. Previous studies have looked at specific mechanisms of CI in the digital context, but by deeply interrogating multiple mechanisms of CI for a single case study of Wikipedia, this research suggests a more rounded approach is necessary to accurately translate previous models to the digital media world. The CI on Wikipedia matches similar characteristics of CI in biological and mathematical settings, but at the same time, it also transcends previous conceptions, as the masses of individual actors in the system exhibit intelligence beyond what is necessary for stigmergy and emergence. As such, this research contributes to and expands upon the theoretical work established by Levy (1997) on the human implications of sociotechnical CI.

Within the growing field of wiki and Wikipedia studies, this research serves a number of important functions. First and foremost, this work gives voice to the bot community on Wikipedia. Though a relatively small group of contributors to the site, bot operators and programmers maintain a significant role in the functioning and improvement of the project. Beyond their personal editing, which many times is extensive, their automated creations roam across much of the encyclopedia, protecting against vandalism, enforcing protocol, and welcoming new users, to name a few of their tasks. Overall, bots make nearly 22.5% of all edits to Wikipedia (Zachte, 2011), and yet previous research on bots and the bot community has used aggregated statistical data and content analysis to understand this activity. By interviewing over 40 members of this community, including many who have actively shaped the development of bots and bot policy over the project's short history, this study was able to gain insight into what motivates these Wikipedians, how they view their work, and how they view the site as a whole.

A major finding from this interview research is that many bot operators do not maintain regular and close communication with MediaWiki developers and the

Wikimedia Foundation, a fact confirmed through interviews with key outreach coordinators from the parent organization. Though some on each end feel improved communication among these groups is not a priority (and in many ways a herculean task, as is any communication with a dispersed, open-source community of contributors), various frustrations with MediaWiki and the technical configurations of the site emerged from these interviews. Many bots are created to handle tasks not natively managed by MediaWiki, and while this does not signal a need for MediaWiki developers to integrate all of these functions—remember, MediaWiki is purposefully kept light and flexible—it does indicate the importance of collaboration and feedback between the groups, as bots occupy an important position in Wikipedia's ecosystem. Additionally, with the plateau and possible decline of new contributors to the Wikipedia as a whole, it would be prudent of the WMF to strengthen its ties with the bot community, a group of invested contributors largely preventing the degradation of the project.

Limitations and Future Research

As discussed in Chapter III, this study was bound by certain limitations arising from its methodology and scope. In acknowledging and considering these limitations, other important future lines of research on Wikipedia specifically and digital media more broadly become apparent.

This study looked at both the technology and the community that make Wikipedia work, but within this dynamic ecosystem, the research focused in on the roles and functions bots and bot operators. This focus was purposive, as previous research (Geiger & Ribes, 2010; Geiger, 2011; Niederer & van Dijck, 2010) suggested these groups of actors occupy a uniquely important position in the network of actors, a fact that this study confirmed. However, other groups play an important part in propagating this sociotechnical system, particularly MediaWiki programmers and developers. Addition research should consider in more detail the work of this population, investigating their relationships to the technology and the community of the project. Of particular interest may be MediaWiki developers' ties to the bot community, as there is important overlap[67] in the work and concerns of these two groups.

Additionally, unlike the bot community, who continue to offer their time, effort, and expertise to the project on a completely voluntary basis, the MediaWiki developer community has become fractured over time. Primary development of the MediaWiki core code has been centralized by the Wikimedia Foundation to its paid engineering staff, though an unknown but likely large number of volunteers continue to contribute from the fringes. This makes for a hybrid workflow marked by both collaboration and conflict, as motivations to participate can vary greatly between these two groups. Further work could explore the implications of the MediaWiki development process from sociological and political economic perspectives, teasing out how power is aggregated and exercised in this arrangement.

Expanding the sample of Wikipedia activity considered in this study's exploratory network analysis would also advance our understanding of sociotechnical actors in the network of collaboration. Kane (2009) indicates a connection between patterns of collaboration and article quality on the site, showing the number of editors and the average edits per editor are significantly correlated to article quality and article traffic, though his work does not consider bots in the milieu of contributors. Building from these

[67] There is literal overlap, as well; seven of the bot operators interviewed for this study are also MediaWiki developers.

findings and the present research's suggestion that bots are central in many ways to this collaboration, a comprehensive network analysis could reveal more precisely how heterogeneous actors work together to create quality and lasting content. In addition, a more complete sample of editing histories could explore the specific nature of each actor's work, revealing who contributes major edits, minor edits, vandalism protection, formatting labor, and other types of work. Detailed qualitative research could follow to further frame the various meanings that contributors give to creating content.

More broadly, this research used Wikipedia as a case study to examine how theories of sociotechnical systems explain, clarify, and illuminate the activity of digital media participation. Findings here indicate the actors and structures on Wikipedia are bound together in an ongoing process of development and maintenance enabled by the code-based, immaterial nature of its medium, from which a collective cyberculture emerges. This case study suggests that similar research into other digital media platforms is warranted and could heighten our understanding of the digital world. A number of user-generated content sites share fundamental similarities with Wikipedia, including a knowledge or education-based mission and a large contributor population with a smaller core of invested users, yet have been vastly understudied to this point. WordPress and Urban Dictionary are but two examples. WordPress, a popular blogging website, is a free and open-source software project that is programmed and developed by users while being supported by a parent non-profit foundation. Both its content and code are dynamic, yet stable enough to power nearly 15% of the globe's websites (Rao, 2011). Urban Dictionary, a more centralized, commercialized site, features over 6.5 million definitions submitted by users and vetted by a volunteer editorial staff. By studying these projects and others that fall along the spectrum of open-source/commercial, we can more thoroughly and intricately understand the sociotechnical and socioeconomic implications of the Web 2.0 paradigm of participation as it forms the foundation for the next stage of the Internet and digital media.

Of Bots and Black Boxes: Concluding Thoughts

Most people experience Wikipedia as a black box, which is to say they use the world's largest encyclopedia as a solid and reliable reference tool. Some come across it unintentionally, perhaps guided by Google, as many articles appear at the top of the search engine's results, or even by a librarian, many of whom now recommend the site as a jumping off point for research. Others use the site much more purposefully; as a physician recently told this researcher, he keeps a shortcut to Wikipedia on his computer's desktop, not to verify a diagnosis in the traditional sense, but to just "double check" he has his facts straight. Indeed, confidence in Wikipedia as an accurate and up-to-date source of information has grown tremendously over the past decade, buoyed by media coverage, academic research, but perhaps most importantly, everyday experience.

Most people experience Wikipedia as a black box, a largely finished product, though Talk, Edit, and History pages beckon the reader of each and every article to open that box and explore the social and technical collaboration behind the information they consume. Those who do are exposed to a dynamic system of actors, structures, protocols, and bureaucracy—a sociotechnical system maintained not by a central authority, but by the constant interactions of these elements in a flow of activity. Like the very nature of digital media, which needs to be constantly rendered to be made useful and material, Wikipedia itself needs constant internal movement to appear outwardly still.

This study explored this apparent paradox of the black box, dissecting the movement within to understand how contributors, both human and technological, create a stable site used by millions of Internet travelers each day. Like Marx's commodity form, the black box is a theoretical construct, offered by actor-network theory to understand the complex, fluctuating, and often hidden relationships between social relations and the material world. As the online world continues to grow in importance for our daily lives, we must remain diligent in our attention to the black boxes that spring up across the digital landscape … who knows what hybrids may lay inside.

APPENDIX A

INFORMED CONSENT LETTER

Greetings Wikipedia Editor-

My name is Randall Livingstone, and I am a doctoral student at the University of Oregon, studying digital media and online communities. I am posting to invite you to participate in my dissertation research exploring the work of Wikipedia editors, programmers, and administrators involved in the creation and management of bots and automated editing tools. The interview should take 30 to 45 minutes and can be conducted online or in-person, at your convenience.

Your interview responses will help online communication researchers like me to better understand the collaborations, challenges, and purposeful work of Wikipedia editors and programmers like you. Questions will address your contributions to Wikipedia, and specifically, your thoughts, opinions and experiences working with bots, automated editing tools, and technical protocol in the system.

The interview is voluntary, and your confidentiality will be protected. You will have the choice of using your real name, Wikipedia username, or creating a unique pseudonym during the research. Your name will not appear on any documents or in the actual dissertation if you wish to keep it confidential. If you choose to keep your name secret, I will be the only one with access to materials including your name. If you agree to be interviewed online and would not like your name to appear in the dissertation, please note that online communication is not always completely secure and absolute privacy is not guaranteed. I will do my best to protect your information by immediately removing your name from saved text files, and all information will be encrypted and save only on my personal computer immediately following the interview.

There are no foreseeable risks to you by participating in this study. You may choose to not answer any question or questions, and you may withdraw at any time. By agreeing to be interviewed, you are providing informed consent to participate in the research and for the interview to be recorded via audio or video (when applicable). You may also obtain a copy of this consent form at any time.

If you have any questions about the study, please contact me via my Talk Page (UOJComm) or at livingst@uoregon.edu. My faculty advisor is Dr. Bish Sen, who may be reached at bsen@uoregon.edu. If you have any questions regarding your rights as a research participant, please contact the Office for Protection of Human Subjects at human_subjects@orc.uoregon.edu.

Thank you very much.

Sincerely,

Randall Livingstone
University of Oregon
School of Journalism & Communication
1275 University of Oregon
Eugene, OR 97403

APPENDIX B

PAGES EXAMINED FOR DOCUMENT ANALYSIS

Phase 1 (listed alphabetically)

Manual:FAQ

Manual:What is MediaWiki?

MediaWiki (article)
Talk:MediaWiki (November 2004 through July 2011)

Wikipedia:Bot Approvals Group
Wikipedia talk:Bot Approvals Group (March 2006 through December 2011)

Wikipedia:Bot Policy
Wikipedia talk:Bot Policy/Archives 1-23 (mid-2002 through December 2011)

Wikipedia:Bots
Wikipedia talk:Bots (November 2009 through October 2011)

Wikipedia:Bots/Requests for approval
Wikipedia talk:Bots/Requests for approval (March 2006 through December 2011)

Wikipedia:Creating a bot
Wikipedia talk:Creating a bot (December 2006 through December 2011)

Wikipedia:History of Wikipedia bots
Wikipedia talk:History of Wikipedia bots (late-2002 through October 2011)

Wikipedia:Types of bots

APPENDIX C

INTERVIEW INVITATION MESSAGE

Greetings-

My name is Randall Livingstone, and I am a graduate student at the University of Oregon, currently collecting data for my dissertation on Wikipedia editors who create and use bots and assisted editing tools, as well as editors involved in the initial and/or ongoing creation of bot policies on Wikipedia. As a member of the bot community and bot operator, I would very much like to interview you for the project at a time and in a method that is most convenient for you (Gchat, another IM client, Skype, email, telephone, etc.). I am completely flexible and can work with your schedule. The interview will take approximately 45-60 minutes.

My dissertation project has been approved both by the [http://humansubjects.uoregon.edu/ Institutional Review Board (IRB)] at the University of Oregon, and by the [http://meta.wikimedia.org/wiki/Research:Committee Research Committee] at the Wikimedia Foundation. You can find more information on the project on my [http://meta.wikimedia.org/wiki/Research:Understanding_the_Editor/Bot_Relationship meta page].

Please let me know if you have any questions, and I look forward to hearing from you to set up a time to chat. Thank you very much.

Randall Livingstone, School of Journalism & Communication, University of Oregon

~~~~ [Wikipedia signature]

<u>Posted to the following Wikipedia pages:</u>

Wikipedia talk:Bot Approvals Group (January 5, 2011)
Wikipedia:Bot owners' noticeboard (January 25, 2011)
Wikipedia talk:Bot Policy (January 5, 2011)
Wikipedia:Village pump (technical) (January 5, 2011)

<u>Sent to the following Wikipedia mailing list:</u>

pywikipedia-l (February 20, 2011)
~~~~

APPENDIX D

INTERVIEW GUIDE

Are you over the age of 18? (I have to ask this to start, as I am not approved to interview anyone under that age.)

General:

It looks like you started contributing to the English WP from the [username] account in 20XX. Where these your first contributions to WP as a whole? Why did you start contributing to the project as a whole? What were your motivations at the time? Was there a particular thing that got you involved?

You became an Admin in [month] 20XX. Do you have any other user privileges on the English WP? Other versions of WP?

What (if any) semi-automated tools or assisted-editing tools do you use with WP?

Could you tell me a bit more about your previous experience as a programmer, both professionally and personally (if any)?

Both on WP and in general, what programming languages and/or platforms do you program with (if any)?

Please provide any demographic data about yourself that you feel comfortable providing:

- Gender
- Age
- Education level (if college, what subject(s) is/are your degree(s) in?)
- Current city and nation of residence
- Birth city and nation
- Languages you are fluent in

User's Bot:

You are the operator of [bot username], which was approved in 20XX for use on the English WP.

- Why did you create the bot?
- In lay person's terms, what does your bot do? (if you can sum it up)
- What other tasks (if any) does the bot do? Were they added after the initial approval?
- Does the bot run on the Toolserver? If not, where does it run?
- Does the edit count matter to you? What does it represent to you?

Do you remember much about the approval process for your bot? What was that experience like (beyond what is documented in the BRFA)?

Have you ever experienced a conflict with another editor regarding your bot? Please describe.

Have you programmed or operated any other bots? If so, for what reasons did you create the bot(s)?

Bot Community:

When did you become involved with the WP bot community? Why did you become involved in the bot community? What motivated you? Was there a particular task, incident, fellow editor, etc. that spurred you?

What have your experiences [with/on] the Bots Approval Group (BAG) been like? Please describe your experiences.

What other WP bot operators do you/have you interact(ed) with on a regular basis? Please list as many as you can think of.

Bots – General:

Do you feel bots and automated tools are necessary? Beneficial? Harmful?

Have you ever seen a bot “go bad” and cause harm to WP? If so, what happened? Were you involved in discussions around the incident? Did you take any actions?

Have you been active at all in the creation or revision of Bot Policy?

Do you feel the Wikipedia community (beyond those immediately involved with bots) is supportive of bots and automated tools on WP?

Do you feel non-editing users (i.e. readers) of Wikipedia generally know about the work of bots and the editors who create them? The amount of work that bots do to maintain the site? Do you think it’s necessary or important for them to know about your work?

Are there tasks you think bots shouldn’t be allowed to do? Why? (for example, some editors don’t think bots should deal with spelling issues.)

Are there tasks you think human editors shouldn’t be allowed to do (only bots)?

Wikipedia:

Some popular press accounts of Wikipedia have described the project as an instance of collective intelligence. How would you define collective intelligence? Do you feel WP is an example of collective intelligence?

Presently, what are your motivations for contributing automated tools to Wikipedia? What do you feel you personally get from this work? Have the motivations and benefits changed since you’ve joined the site?

What do you feel the strengths and weaknesses are of Wikipedia and its technical structure? Its administrative structure?

How do you feel about the sanction systems (like ArbCom) on Wikipedia for editors? The recognition systems (like Barnstars)?

Would you improve anything about Wikipedia's governance? If so, what?

Do you have any other comments, experiences, anecdotes, or stories you'd like to share that might help my project, which endeavors to understand and give voice to Wikipedia bots and their creators?

Overall, how would you describe your interactions with other Wikipedia editors? Other administrators?

Are there any particular content topics that you focus on in your main namespace editing contributions?

APPENDIX E

DATA ANALYSIS CATEGORIES AND SUBCATEGORIES

Social
- Edit count
- Roles
- Labor
- Bureaucracy
- Article / code ownership and development
- New Users
- Credibility

Tasks and permissions
- Adminbots
- Bot vs. human tasks
- Bots gone bad
- Bot tasks – vandalism, copyright

Collaboration and conflict
- Communication
- Specific collaborations
- Specific conflicts
- Bots left behind

Technical Performance
- Bot accounts
- Toolserver

MediaWiki

Wikimedia Foundation
- Operations
- WMF / bot operator communication

Bots (general)
- Beneficial / Necessary / Harmful?
- Getting involved in the bot community

Bot policy

- Policies
- Bot speed

Bot Approvals Group

Global bots

Community reaction to bots

- Wikipedia community
- General readers
- Technophobia / technophilia

Semi-automated tools

Collective Intelligence

REFERENCES CITED

Abbate, J. (1999). *Inventing the Internet*. Cambridge, MA: MIT Press.

Adams, C. (2001, January 27). Hello, "FreeLinks" work, and redirected pages. Retrieved from http://lists.wikimedia.org/pipermail/wikipedia-l/2001-January/000007.html

Adams, C. (2007). UseModWiki. Retrieved from http://www.usemod.com/cgi-bin/wiki.pl

Andrejevic, M. (2002). The work of being watched. *Critical Studies in Media Communication*, *19*(2), 230-248.

Angwin, J., & Fowler, G. A. (2009, November 27). Volunteers log off as Wikipedia ages. *The Wall Street Journal*. Retrieved from http://online.wsj.com/article/SB125893981183759969.html

Applegate, D. L., Bixby, R. M., Chvatal, V, & Cook, W. J. (2006). *The traveling salesman problem*. Princeton, NJ: Princeton University Press.

Ayers, P., Matthews, C., & Yates, B. (2008). *How Wikipedia works: And how you can be a part of it*. San Francisco, CA: No Starch Press.

Bagdikian, B. H. (2004). *The new media monopoly*. Boston, MA: Beacon Press.

Barabási, A.-L. (2002). *Linked: The new science of networks*. Cambridge, MA: Perseus Pub.

Barbaro, M. (2006, March 7). Wal-Mart enlists bloggers in P.R. campaign. *The New York Times*. Retrieved from http://www.nytimes.com/2006/03/07/technology/07blog.html

Baytiyeh, H., & Pfaffman, J. (2010). Volunteers in Wikipedia: Why the community matters. *Journal of Educational Technology & Society*, *13*(2), 128-140.

Bell, D. (1973). *The coming of post-industrial society: A venture in social forecasting*. New York, NY: Basic Books.

Beniger, J. R. (1986). *The control revolution: Technological and economic origins of the information society*. Cambridge, MA: Harvard University Press.

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.

Benoit-Barne, C. (May 01, 2007). Socio-technical deliberation about free and open source software: Accounting for the status of artifacts in public life. *Quarterly Journal of Speech*, *93*(2), 211-235.

Berg, B. (2007). *Qualitative research methods for the social sciences*. Boston, MA: Pearson.

Bergsma, M. (2007). Wikimedia architecture. Retrieved from http://www.nedworks.org/~mark/presentations/kennisnet/Wikimedia%20architecture%20%28kennisnet%29.pdf

Berners-Lee, T. (1998, May 7). The World Wide Web: A very short personal history. Retrieved from http://www.w3.org/People/Berners-Lee/ShortHistory.html

Berners-Lee, T., & Fischetti, M. (1999). *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. San Francisco, CA: HarperSanFrancisco.

Best, K. (2009). When mobiles go media: Relational affordances and present-to-hand digital devices. *Canadian Journal of Communication, 34,* 397-414.

Bijker, W. E. (1995a). *Of bicycles, bakelites, and bulbs: Toward a theory of sociotechnical change*. Cambridge, MA: MIT Press.

Bijker, W. E. (1995b). Sociohistorical technology studies. In S. Jasanoff, G. E. Markle, J. C. Petersen, & T. Pinch (Eds.), *Handbook of science and technology studies* (pp. 229-256). Thousand Oaks, CA: Sage.

Bijker, W. E., & Law, J. (1992). *Shaping technology/building society: Studies in sociotechnical change*. Cambridge, MA: MIT Press.

Bilgil, M. (Director). (2009). *History of the Internet*. [Motion picture]. Retrieved from http://www.youtube.com/watch?v=9hIQjrMHTv4

Binet, A. & Simon, T. (1980). *The development of intelligence in children*. Nashville, TN: Williams Printing Co.

Blodget, H. (2009, January 3). Who the hell writes Wikipedia anyway? *Business Insider*. Retrieved from http://articles.businessinsider.com/2009-01-03/tech/29988468_1_jimmy-wales-wikipedia-insiders

Boerner, G. (2010). Prof. Boerner's explorations. Retrieved from http://www.boerner.net/jboerner/?p=7540

Boltanski, L., & Chiapello, E. (2005). *The new spirit of capitalism*. London, England: Verso.

Bolter, J. D. & Grusin, R. (1996). Remediation. *Configurations, 4*(3), 311-358.

Bolter, J. D. & Grusin, R. (2000). *Remediation: Understanding new media*. Cambridge, MA: MIT Press.

Bonabeau, E. (1999). Editor's introduction: Stigmergy. *Artificial Life, 5*(2), 95-96.

Bongwon, S., Convertino, G., Chi, E. H., & Pirolli, P. (2009). The singularity is not near: Slowing growth of Wikipedia. *Proceedings from WikiSym '09*. Orlando, FL.

Bosman, J. (2012, March 13). After 244 years, Encyclopaedia Britannica stops the presses. *The New York Times*. Retrieved from http://mediadecoder.blogs.nytimes.com/2012/03/13/after-244-years-encyclopaedia-britannica-stops-the-presses/#

Bowden, G. (1995). Coming of age in STS: Some methodological musings. In S. Jasanoff, G. E. Markle, J. C. Petersen, & T. Pinch (Eds.), *Handbook of science and technology studies* (pp. 64-79). Thousand Oaks, CA: Sage.

Brabham, D. (2010). Moving the crowd at threadless: Motivations for participation in a crowdsourcing application. *Information, Communication, and Society*, *13*(8), 1122-1145.

Bradley, G. (2006). *Social and community informatics: Humans on the net*. London, England: Routledge.

Braverman, H. (1998). *Labor and monopoly capital: The degradation of work in the twentieth century*. New York, NY: Monthly Review Press.

Bruns, A. (2007). About Axel Bruns. Retrieved from http://produsage.org/about

Bruns, A. (2008). *Blogs, Wikipedia, Second life, and beyond: From production to produsage*. New York, NY: Peter Lang.

Bryl, V., Giorgini, P., & Mylopoulos, J. (2009). Designing socio-technical systems: from stakeholder goals to social networks. *Requirements Engineering, 14*(1), 47-70.

Burback, R. L. (1998). Abstraction. Retrieved from http://infolab.stanford.edu/~burback/watersluice/node147.html

Butler, B., Joyce, E., & Pike, J. (2008). Don't look now, but we've created a bureaucracy: The nature and roles of policies and rules in Wikipedia. *Proceeding of the Twenty-sixth Annual SIGCHI Conference on Human Factors in Computing Systems* (p. 1101–1110).

Calabresi, M. (2009, April 8). Wikipedia for spies: The CIA discovers Web 2.0. *Time*. Retrieved from http://www.time.com/time/nation/article/0,8599,1890084,00.html

Calacanis, J. (2006, October 28). Wikipedia leaves $100M on the table (or "PLEASE Jimbo, reconsider-media philanthropy could change the world!"). Retrieved from http://calacanis.com/2006/10/28/wikipedia-leaves-100m-on-the-table-or-please-jimbo-reconsider/

Callon, M. (1986a). The sociology of an actor-network: The case of the electric vehicle. In M. Callon, J. Law, & A. Rip (Eds.), *Mapping the dynamics of science and technology* (pp. 19-34). London, England: Macmillan.

Callon, M. (1986b). Some elements of a sociology of translation: Domestication of the scallops and the fisherman of St. Brieuc Bay. In J. Law (Ed.), *Power, action and belief: A new sociology of knowledge* (pp. 196-233). London, England: Routledge & Kegan Pual.

Campbell-Kelly, M. (2003). *From airline reservations to Sonic the Hedgehog: A history of the software industry*. Cambridge, MA: MIT Press.

Carr, N. (2005, October 18). The law of the wiki. *Rough Type.com*. Retrieved from http://www.roughtype.com/archives/2005/10/the_law_of_the.php

Carrington, P. J., Scott, J., & Wasserman, S. (2005). *Models and methods in social network analysis*. Cambridge, England: Cambridge University Press.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. Cambridge, England: Cambridge University Press.

Castells, M. (1996). *The rise of the network society*. Malden, MA: Blackwell Publishers.

Castells, M. (2001). *The Internet galaxy: Reflections on the Internet, business, and society*. Oxford, England: Oxford University Press.

Castells, M. (2004). *The network society*. Cheltenham, England: Edward Elgar.

Center for Collective Intelligence. (2011a). Home. Retrieved from http://cci.mit.edu/index.html

Center for Collective Intelligence. (2011b). What is collective intelligence and what will we do about it? Retrieved from http://cci.mit.edu/about/MaloneLaunchRemarks.html

Ceruzzi, P. E. (1998). *A history of modern computing*. Cambridge, MA: MIT Press.

Chalmers, D. J. (2002). Strong and weak emergence. Retrieved from http://consc.net/papers/emergence.pdf

Cheshire, C. (2010). None of us is as lazy as all of us: Social intelligence and loafing in information pools. *Information Communication & Society*, *13*(4), 537-555.

Choi, S. Y., Kang, Y. S., & Lee, H. (2008). The effects of socio-technical enablers on knowledge sharing: An exploratory examination. *Journal of Information Science, 34*(5), 742-754.

Christakis, N. A., & Fowler, J. H. (2009). *Connected: The surprising power of our social networks and how they shape our lives*. New York, NY: Little, Brown and Co.

Clark, A. (2003). *Natural-born cyborgs: Minds, technologies and the future of human intelligence*. Oxford, England: Oxford University Press.

Colbert, S., & Hoskinson, J. (2007). *The Colbert Report* [Television series]. New York, NY: Comedy Central.

Collison, R. L. (1964). *Encyclopaedias: their history throughout the ages*. New York, NY: Hafner Pub. Co.

Computer History Museum. (2006). An evening with wiki inventor Ward Cunningham in conversation with John Gage. Retrieved from http://video.google.com/videoplay?docid=-7739076742312910146

Computer History Museum. (2012). Making sense of the census: Hollerith's punched card solution. Retrieved from http://www.computerhistory.org/revolution/punched-cards/2/2

Corning, P. A. (2002). The re-emergence of "emergence": A venerable concept in search of a theory. *Complexity, 7*(6), 18-30.

Couldry, N. (2004). Actor-network theory and media: Do they connect and on what terms? In A. Hepp, F. Krotz, S. Moores, & C. Winter (Eds.), *Connectivity, networks, and flows: Conceptualizing contemporary communications* (pp. 93-110). Cresskill, NJ: Hampton Press.

Coupland, D. (1995). *Microserfs*. New York, NY: ReganBooks.

Cunningham, W. (2002). What is wiki. Retrieved from http://www.wiki.org/wiki.cgi?WhatIsWiki

Cunningham, W. (2003, November). Correspondence on the etymology of wiki. Retrieved from http://c2.com/doc/etymology.html

Cypher, M. & Richardson, I. (2006). An actor-network approach to games and virtual environments. *Proceedings from the 2006 International Conference on Game Research and Development*. Perth, Australia: Murdoch University.

Danese. (2011, March 7). Brion Vibber rejoins Wikimedia Foundation. *Wikimedia blog*. Retrieved from http://blog.wikimedia.org/tag/brion-vibber/

Day, R. E. (2001). *The modern invention of information: Discourse, history, and power*. Carbondale, IL: Southern Illinois University Press.

De Vreede, J. (2008). Board of trustees/Restructure announcement. Retrieved from http://wikimediafoundation.org/w/index.php?title=Board_of_Trustees/Restructure_Announcement&oldid=26599

Deary, I. J. (2001). *Intelligence: A very short introduction*. Oxford, England: Oxford University Press.

Dellarocas, C., & Wood, C. A. (2008). The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science, 54*(3), 460-476.

Deuze, M. (2006). Participation, remediation, bricolage: Considering principal components of a digital culture. *The Information Society, 22*, 63-75.

Digital Methods Initiative. (2012). Digital methods. Retrieved from https://www.digitalmethods.net/Digitalmethods/WebHome

Dijk, J. V. (1999). *The network society: Social aspects of new media*. London, England: Sage Publications.

DiNucci, D. (1999). Fragmented future. *Print, 53*(4), 32-33.

Dodds, P. S., Muhamad, R., & Watts, D. J. (2003). An experimental study of search in global social networks. *Science, 301*(5634), 827-829.

Ebersbach, A., Glaser, M., Heigl, R., & Dueck, G. (2006). *Wiki: Web collaboration*. Berlin, Germany: Springer.

Edwards, P. N. (2003). Infrastructure and modernity: Force, time, and social organization in the history of sociotechnical systems. In T. J. Misa, P. Brey, & A. Feenberg (Eds.), *Modernity and technology* (pp. 185-226): Cambridge, MA: MIT Press.

Edwards, P. N. (2010). Some say the Internet should never have happened. In W. R. Neuman (Ed.), *Media, technology, and society: Theories of media evolution* (pp. 141-160): Ann Arbor, MI: University of Michigan Press.

Elgali, Z. & Kalman, Y. M. (2010). Using actor-network theory to analyze the construction of the failure concept in a K-12 ICT integration program. *Proceedings from the Chais Conference on Instructional Technologies Research 2010: Learning in the Technological Era*. Raanana, IL: The Open University of Israel.

Elliott, M. (2006). Stigmergic collaboration: The evolution of group work. *M/C Journal, 9*(2). Retrieved from http://journal.media-culture.org.au/0605/03-elliott.php

Emery, F. E., & Trist, E. L. (1965). The causal texture of organizational environments. In W. A. Pasmore & J. J. Sherwood (Eds.), *Sociotechnical systems: A sourcebook* (pp. 13-27). San Diego, CA: University Associates, Inc.

Essam, M. (2012). The role of social networking sites (SNSs) in the January 25th Revolution in Egypt. *Library Review*, *61*(2) 128-159.

Fisher, E. (2010). *Media and new capitalism in the digital age: The spirit of networks*. New York, NY: Palgrave Macmillan.

Fisher, L. (2009). *The perfect swarm: The science of complexity in everyday life*. New York, NY: Basic Books.

Flew, T. (2002). *New media: An introduction*. South Melbourne, Australia: Oxford University Press.

Fontana, A. & Frey, J. H. (2005). The interview: From neutral stance to political involvement. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (pp. 695-728). Thousand Oaks, CA: Sage.

Forte, A., Larco, V., & Bruckman, A. (2009). Decentralization in Wikipedia governance. *Journal of Management Information Systems*, *26*(1), 49-72.

Fowler, R. (1996). Encyclopaedias: Definitions and theoretical problems. In P. Binkley (Ed.), *Pre-modern encyclopaedic texts* (pp. 3-30). Leiden, The Netherlands: Brill.

Freemon, L. C. (2005). Graphic techniques for exploring social network data. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 248-269). Cambridge, UK: Cambridge University Press.

Fuchs, C. (2008). *Internet and society: Social theory in the information age*. New York, NY: Routledge.

Fuchs, C. (2011a). *Foundations of critical media and information studies*. Abingdon, England: Routledge.

Fuchs, C. (2011b). Web 2.0, prosumption, and surveillance. *Surveillance & Society*, *8*(3), 288-309.

Fuchs, C. (2012). The political economy of Facebook. *Television and New Media*, *13*(2), 139-159.

Fuchs, C., Boersma, K., Albrechtslund, A., & Sandoval, M. (2011). *The Internet and surveillance*. New York, NY: Routledge.

Fuller, S. (2006). *The philosophy of science and technology studies*. New York, NY: Routledge.

Galloway, A. R. (2004). *Protocol: How control exists after decentralization*. Cambridge, MA: MIT Press.

Galloway, A. R., & Thacker, E. (2007). *The exploit: A theory of networks*. Minneapolis, MN: University of Minnesota Press.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York, NY: Basic Books.

Gardner, S. (2011). Sue Gardner's blog: About. Retrieved from http://suegardner.org/about/

Garton Ash, T. (2011, January 12). We've seen America's vitriol. Now let's salute Wikipedia, a US pioneer of global civility. *The Guardian*. Retrieved from http://www.guardian.co.uk/commentisfree/cifamerica/2011/jan/12/wikipedia-us-pioneer-global-idealism?INTCMP=SRCH

Geiger, R. S. (2011). The lives of bots. In G. Lovink & N. Tkacz (Eds.), *Critical point of view:A Wikipedia reader* (pp. 78-93). Amsterdam, NL: Institute of Network Cultures.

Geiger, R. S., & Ribes, D. (2010). The work of sustaining order in Wikipedia: The banning of a vandal. *Proceedings from the 2010 ACM Conference on Computer Supported Cooperative Work*. Savannah, GA.

Gillespie, T. (2011). Can an algorithm be wrong? Twitter Trends, the specter of censorship, and our faith in the algorithms around us. *CultureDigitally.org*. Retrieved from http://culturedigitally.org/2011/10/can-an-algorithm-be-wrong/

Gleick, J. (2011). *The information: A history, a theory, a flood*. New York, NY: Pantheon Books.

Goertzel, B. (2002). *Creating Internet intelligence: Wild computing, distributed digital consciousness, and the emerging global brain*. New York, NY: Kluwer Academic / Plenum Publishers.

Goldman, E. (2009). Wikipedia's labor squeeze and its consequences. *Journal of Telecommunications and High Technology Law, 8*, 158-184.

Goldspink, C. (2010). Normative behaviour in Wikipedia. *Information, Communication & Society*, *13*(5), 652-673.

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24*(1), 13-23.

Gouthro, L. (2000, March 10). Building the world's biggest encyclopedia. *PCWorld*. Retrieved from http://www.pcworld.com/article/15676/ building_the_worlds_biggest_encyclopedia.html

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360-1380.

Grasse, P. P. (1959). La reconstruction du nid et les coordinations interindividuelles chez Bellicositermes natalensis et Cubitermes sp. La theorie de la stigmergie: Essai d'interpretation du comportement des termites constructeurs. *Insectes Sociaux, 6*(1), 41–83.

Gregoriades, A., & Sutcliffe, A. (2008). A socio-technical approach to business process simulation. *Decision Support Systems, 45*(4), 1017-1030.

Gregorio, J. (2003). Stigmergy and the World-Wide Web. Retrieved from http://bitworking.org/news/Stigmergy

Grigas, V. & Roth, M. (2011). Wikimedia strategic plan. Retrieved from http://vimeo.com/28437347

Hacker, D. & Renshaw, B. (1985). *Writing with a voice: A rhetoric and handbook*. Boston, MA: Little Brown.

Hafner, K., & Lyon, M. (1996). *Where wizards stay up late: The origins of the Internet*. New York, NY: Simon & Schuster.

Halacy, D. S. (1970). *Charles Babbage, father of the computer*. New York, NY: Crowell-Collier Press.

Halavais, A., & Lackaff, D. (2008). An analysis of topical coverage of Wikipedia. *Journal of Computer-Mediated Communication, 13*(2), 429-440.

Hally, M. (2005). *Electronic brains: Stories from the dawn of the computer age*. Washington, D.C: Joseph Henry Press.

Hanneman R. A. & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of California, Riverside. Retrieved from http://faculty.ucr.edu/~hanneman/nettext/

Hardt, M., & Negri, A. (2004). *Multitude: War and democracy in the age of Empire*. New York, NY: Penguin Books.

Hawkins, J., & Blakeslee, S. (2004). *On intelligence*. New York, NY: Times Books.

Helm, B. (2005, December 14). Wikipedia: "A work in progress." *BusinessWeek*. Retrieved from http://www.businessweek.com/technology/content/dec2005/tc20051214_441708.htm?chan=db

Herman, E. S., & Chomsky, N. (1988). *Manufacturing consent: The political economy of the mass media*. New York, NY: Pantheon Books.

Herrnstein, R. J., & Murray, C. A. (1994). *The bell curve: Intelligence and class structure in American life*. New York, NY: Free Press.

Hertel, G., Niedner, S., & Herrmann, S. (2003). Motivation of software developers in Open Source projects: an Internet-based survey of contributors to the Linux kernel. *Research Policy, 32*(7), 1159.

Heylighen, F. (2008). The emergence of a global brain. In M. Tovey (Ed.), *Collective intelligence: Creating a prosperous world at peace* (pp. 305-314). Oakton, VA: Earth Intelligence Network.

Hillman & Carr (Producers). (2011). False dawn: The Babbage engine. *Computer History Museum*. Retrieved from http://www.computerhistory.org/revolution/calculators/1/51/2205

Hine, C. (2000). *Virtual ethnography*. London, England: Sage.

Holland, J., Hutchins, E., & Kirsh, D. (2000). Distribution cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction, 7*(2), 174-196.

Hopper, K. (2008). Empowering individuals toward collective online production. In M. Tovey (Ed.), *Collective intelligence: Creating a prosperous world at peace* (pp. 245-250). Oakton, VA: Earth Intelligence Network.

Horkheimer, M., & Adorno, T. W. (1947). *Dialectic of enlightenment*. New York, NY: Herder and Herder.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Jackson, M. H., Poole, M. S., & Kuhn, T. (2002). The social construction of technology in studies of the workplace. In L. A. Lievrouw & S. Livingstone (Eds.), *The handbook of new media* (pp. 34-49). London, England: Sage.

Jankowski, N. W. (2002). Creating community with media: History, theories and scientific investigations. In L. A. Lievrouw & S. Livingstone (Eds.), *The handbook of new media* (pp. 34-49). London, England: Sage.

Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. New York, NY: New York University Press.

Jesus, R. (2010). *Cooperation and cognition in Wikipedia articles: A data-driven philosophical and exploratory study*. (Doctoral dissertation). Retrieved from http://www.nbi.ku.dk/english/research/phd_theses/phd_theses_2010/rut_jesus/rut_jesus2.pdf/

Jhally, S. (1990). *The codes of advertising: Fetishism and the political economy of meaning in the consumer society*. New York, NY: Routledge.

Jhally, S. (2006). *The spectacle of accumulation: Essays in culture, media, & politics*. New York, NY: Lang.

Johnson, B. (2009, August 12). Wikipedia approaches its limits. *The Guardian*. Retrieved from http://www.guardian.co.uk/technology/2009/aug/12/wikipedia-deletionist-inclusionist

Johnson, D. G., & Wetmore, J. M. (2008). STS and ethics: Implications for engineering ethics. In E. J. Hackett, O. Amsterdamska, M. Lynch, & J. Wajcman (Eds.), *The handbook of science and technology studies* (pp. 567-582). Cambridge, MA: MIT Press.

Johnson, N. L. (2008). Science of CI: Resources for change. In M. Tovey (Ed.), *Collective intelligence: Creating a prosperous world at peace* (pp. 265-274). Oakton, VA: Earth Intelligence Network.

Johnson, S. (2001). *Emergence: The connected lives of ants, brains, cities, and software*. New York, NY: Scribner.

Jupp, V. (Ed.). (2006). *The Sage dictionary of social research methods*. London, England: Sage.

Kanalley, C. (2011, January 27). Egypt's Internet shut down, according to reports. *The Huffington Post*. Retrieved from http://www.huffingtonpost.com/2011/01/27/egypt-internet-goes-down-_n_815156.html

Kane, G. C. (2009). It's a network, not an encyclopedia: A social network perspective on Wikipedia collaboration. *Proceedings from the 2009 Academy of Management Annual Meeting*, Chicago, IL.

Kane, G. C., Majchrzak, A., Johnson, J., & Chenisern, L. (2009). A longitudinal model of perspective making and perspective taking within fluid online collectives. *Proceedings from the 2009 International Conference on Information Systems*, Phoenix, AZ.

Kane, P. (2012, January 20). SOPA, PIPA votes to be delayed in House and Senate. *The Washington Post*. Retrieved from http://www.washingtonpost.com/blogs/2chambers/post/sopa-senate-vote-to-be-delayed-reid-announces/2012/01/20/gIQApRWVDQ_blog.html

Karbasfrooshan, A. (2006). What is Wikipedia.org's valuation? Retrieved from http://www.watchmojo.com/web/blog/?p=626

Kattouw, R. (2011, December 3). Wikimedia infrastructure. Retrieved from http://wikitech.wikimedia.org/index.php?title=File%3AWikimedia_Infrastructure_Utrecht_LUG_2011.pdf&page=1

Keen, A. (2007). *The cult of the amateur: How today's internet is killing our culture*. New York, NY: Doubleday/Currency.

Kellner, D. (2009). Media industries, political economy, and media/cultural studies: An articulation. In J. Holt & A. Perren (Eds.), *Media industries History, theory, and method* (pp. 95-107). Malden, MA: Wiley-Blackwell.

Kelty, C. M. (2008). *Two bits: The cultural significance of free software*. Durham, NC: Duke University Press.

Kimmerle, J., Moskaliuk, J., Harrer, A., & Cress, U. (2010). Visualizing co-evolution of individual and collective knowledge. *Information, Communication & Society*, *13*(8), 1099-1121.

Kirschenbaum, M. G. (2008). *Mechanisms: New media and the forensic imagination*. Cambridge, MA: MIT Press.

Kittur, A., Chi, E. H., & Suh, B. (2009, January 1). What's in Wikipedia? Mapping topics and conflict using socially annotated category structure. *Chi Conference-, 3*, 1509-1512.

Klein, N. (2000). *No logo: Taking aim at the brand bullies*. New York, NY: Picador.

Kline, R. & Pinch, T. (1999). The social construction of technology. In D. MacKenzie & J. Wajcman (Eds.), *The social shaping of technology* (pp. 113-115). Buckingham, England: Open University Press.

Kling, R. (1999). What is social informatics and why does it matter? *DLIB Magazine, 5*(1). Retrieved from http://www.dlib.org/dlib/january99/kling/01kling.html

Kling, R. (2000). Learning about information technologies and social change: The contribution of social informatics. *The Information Society, 16*(3), 217-232.

Konieczny, P. (2009). Governance, organization, and democracy on the Internet: The Iron Law and the evolution of Wikipedia. *Sociological Forum, 24*(1), 162-192.

Kozinets, R. V. (2010). *Netography: Doing ethnographic research online*. Los Angeles, CA: Sage.

Landow, G. P. (2006). *Hypertext 3.0: Critical theory and new media in an era of globalization*. Baltimore, MD: Johns Hopkins University Press.

Lanier, J. (2006, May 30). Digital Maoism: The hazards of the new online collectivism. *Edge*. Retrieved from http://www.edge.org/3rd_culture/lanier06/lanier06_index.html

Lanier, J. (2010). *You are not a gadget: A manifesto*. New York, NY: Alfred A. Knopf.

Laningham, S. (2006, August 22). developerWorks interviews: Tim Berners-Lee. Retrieved from http://www.ibm.com/developerworks/podcast/dwi/cm-int082206txt.html

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

Latour, B. (1991). Technology is society made durable. In J. Law (Ed.), *A sociology of monsters: Essays on power, technology, and domination* (pp. 103-131). London, England: Routledge.

Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies*. Cambridge, MA: Harvard University Press

Latour, B. (2005). *Reassembling the social: An introduction to actor-network theory*. Oxford, England: Oxford University Press.

Law, J. (1987). Technology and heterogeneous engineering: The case of Portuguese expansion. In W. E. Bijker, T. P. Hughes, & T. Pinch (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology* (pp. 17-50). Cambridge, MA: MIT Press.

Law, J. (1992). Notes on the theory of the actor-network: Ordering, strategy and heterogeneity. *Systems Practice*, *5*(4), 379-393.

Lee, M. (2010). A political economic critique of Google Maps and Google Earth. *Information, Communication & Society, 13*(6), 909-928.

Leonard, A. (1997). *Bots: The origin of new species*. San Francisco, CA: Hardwired.

Leonardi, P. M., & Barley, S. R. (2008). Materiality and change: Challenges to building better theory about technology and organizing. *Information and Organization, 18*(3), 159-176.

Lessig, L. (2004). *Free culture: How big media uses technology and the law to lock down culture and control creativity*. New York, NY: Penguin Press.

Lessig, L. (2006). *Code: Version 2.0*. New York, NY: Basic Books.

Leuf, B., & Cunningham, W. (2001). *The Wiki way: Quick collaboration on the Web*. Boston, MA: Addison-Wesley.

Lévy, P. (1997). *Collective intelligence: Mankind's emerging world in cyberspace*. New York, NY: Plenum Trade.

Lévy, P. (2001). *Cyberculture*. Minneapolis, MN: University of Minnesota Press.

Lievrouw, L. A. & Livingstone, S. (2002). *Handbook of new media: Social shaping and consequences of ICTs*. London, England: Sage.

Lih, A. (2009). *The Wikipedia revolution: How a bunch of nobodies created the world's greatest encyclopedia*. New York, NY: Hyperion.

Lindlof, T. R. & Taylor, B. C. (2011). *Qualitative communication research methods* (3rd ed). Los Angeles, CA: Sage.

Livingstone, R. M. (2010). Let's leave the bias to the mainstream media: A Wikipedia community fighting for information neutrality. *M/C Journal*, *13*(6). Retrieved from http://journal.media-culture.org.au/index.php/mcjournal/article/viewArticle/315

Livingstone, S. (2002). Introduction. In L. A. Lievrouw & S. Livingstone (Eds.), *The handbook of New Media* (pp. 17-21). London, England: Sage.

Loebner.net (2012). Home page of the Loebner Prize in artificial intelligence. Retrieved from http://www.loebner.net/Prizef/loebner-prize.html

Lohr, S. (2012, February 11). The age of Big Data. *The New York Times*. Retrieved from http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?_r=1

Luger, G. F. (2005). *Artificial intelligence: Structures and strategies for complex problem solving*. Harlow, England: Addison-Wesley.

Madanapalle, A. (2009, March 8). Top 10 trolls in Internet history. Retrieved from http://www.thinkdigit.com/Internet/Top-10-Trolls-in-Internet-History_3261.html

Majchrzak, A., Wagner, C., & Yates, D. (2006). Corporate wiki users: Results of a survey. Proceedings from *The 2nd International Symposium on Wikis and Open Collaboration*. Odense, Denmark.

Manjoo, F. (2009, September 30). Jimmy Wales quietly edits Wikipedia's new edit policy. *Time*. Retrieved from http://www.time.com/time/business/article/0,8599,1926826,00.html

Manovich, L. (2001). *The language of new media*. Cambridge, MA: MIT Press.

Markham, A. N. & Baym, N. K. (Eds.). (2009). *Internet inquiry: Conversations about method*. Los Angeles, CA: Sage.

Markoff, J. (2011, November 7). It started digital wheels turning. *The New York Times*. Retrieved from http://www.nytimes.com/2011/11/08/science/computer-experts-building-1830s-babbage-analytical-engine.html?_r=2

Marx, K. (1867). *Capital: A critique of political economy*. Retrieved from http://www.marxists.org/archive/marx/works/1867-c1/

Mattelart, A., & Mattelart, M. (1998). *Theories of communication: A short introduction*. London, England: Sage.

Mauldin, M. L. (January 01, 1994). ChatterBots, TinyMuds, and the Turing Test: Entering the Loebner Prize Competition. *Proceedings of the National Conference on Artificial Intelligence*, 16.

Maxwell, J. W., & Felczak, M. (2008). Success through simplicity: On developmental writing and online communities of inquiry. In R. E. Cummings & M. Barton (Eds.), *Wiki writing: Collaborative learning in the college classroom* (pp. 90-108). Ann Arbor, MI: Digital Culture Books.

Mayer, D. (2003, July 19). Phase IV, Wikibooks.org/.com and WikimediaFoundation.org/.com (was wikis and uniformity). Retrieved from http://lists.wikimedia.org/pipermail/wikipedia-l/2003-July/011021.html

McChesney, R. W. (2008). *The political economy of media: Enduring issues, emerging dilemmas*. New York, NY: Monthly Review Press.

McCorduck, P. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. Natick, MA: A.K. Peters.

McQuail, D. (2005). *McQuail's mass communication theory*. London, England: Sage.

MediaWiki. (2012a). Extensions:Gadgets. Retrieved from http://www.mediawiki.org/wiki/Extension:Gadgets

MediaWiki. (2012b). Manual:Extensions. Retrieved from http://www.mediawiki.org/wiki/Manual:Extensions

MediaWiki. (2012c). Manual:What is MediaWiki? Retrieved from http://www.mediawiki.org/wiki/Manual:What_is_MediaWiki%3F

MediaWiki. (2012d). MediaWiki history. Retrieved from http://www.mediawiki.org/wiki/MediaWiki_history

MediaWiki. (2012e). Visual editor. Retrieved from http://www.mediawiki.org/wiki/Visual_editor

MediaWiki. (2012f). Wikimedia engineering report/2012/January. Retrieved from https://www.mediawiki.org/wiki/Wikimedia_engineering_report/2012/January

Meehan, E. R., & Riordan, E. (2002). *Sex & money: Feminism and political economy in the media*. Minneapolis, MN: University of Minnesota Press.

Meehan, E. R. & Torre, P. J. (2011). Markets in theory and markets in television. In J. Wasko, G. Murdock, & I. Sousa (Eds.), *The handbook of political economy of communications* (pp. 62-82). West Sussex, England: Wiley-Blackwell.

Memcached.org. (2012). What is Memcached? Retrieved from http://memcached.org/

Mendelsohn, B. (2011). *Bundled, buried, and behind closed doors*. Retrieved from http://benmendelsohn.com/

Miller, R. (2011). Who has the most Web servers? *Data Center Knowledge*. Retrieved from http://www.datacenterknowledge.com/archives/2009/05/14/whos-got-the-most-web-servers/

Moeller, E. (2007, March 6). Rename admins to janitors. Retrieved from http://marc.info/?l=wikien-l&m=117318790523181

Morell, M. F. (2011). The Wikimedia Foundation and the governance of Wikipedia's infrastructure: Historical trajectories and its hybrid character. In G. Lovink & N. Tkacz (Eds.), *Critical point of view: A Wikipedia reader* (pp. 325-341). Amsterdam, The Netherlands: Institute of Network Cultures.

Moreno, J. L. (1934). *Who shall survive?: Foundations of sociometry, group psychotherapy and sociodrama*. Beacon, NY: Beacon House.

Mosco, V. (2004). *The digital sublime: Myth, power, and cyberspace*. Cambridge, MA: MIT Press.

Mosco, V. (2009). *The political economy of communication*. London, England: SAGE.

Mosco, V. (2011). The political economy of labor. In J. Wasko, G. Murdock, & I. Sousa (Eds.), *The handbook of political economy of communications* (pp. 358-380). West Sussex, England: Wiley-Blackwell.

Murdock, G. (2010, March 30). The return of the gift: Participation and exploitation on the Internet. SOJC Communication and Society Annual Lecture, Eugene, OR.

Murdock, G. & Golding, P. (1973). For a political economy of mass communications. In J. Wasko, G. Murdock, & I. Sousa (Eds.), *The handbook of political economy of communications* (pp. 345-366). West Sussex, England: Wiley-Blackwell.

Negroponte, N. (1995). *Being digital*. New York, NY: Knopf.

Neisser, U., Boodoo, G., Bouchard, T. J. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., ... Urbina, S. (January 01, 1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*(2), 77-101.

Newman, M. E. J. (2010). *Networks: An introduction*. Oxford, England: Oxford University Press.

Niederer, S., & van Dijck, J. (2010). Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. *New Media & Society, 12*(8), 1368-1387.

Nilsson, N. J. (2010). *The quest for artificial intelligence: A history of ideas and achievements*. Cambridge, England: Cambridge University Press.

Norris, P. (2001). *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. Cambridge, England: Cambridge University Press.

Nov, O. (2007, November). What motivates Wikipedians? *Communications of the ACM, 50*, 60–64.

Oboler, A., Steinberg, G., & Stern, R. (2010). The framing of political NGOs in Wikipedia through criticism elimination. *Journal of Information Technology & Politics, 7*, 284-299.

Oreg, S., & Nov, O. (2008). Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values. *Computers in Human Behavior*, *24*(5), 2055–2073.

Ortega, J. F. (2009). *Wikipedia: A quantitative analysis*. Retrieved from http://felipeortega.net/Wikipedia

Osterloh, M. & Rota, S. (2004). Trust and community in open source software production. *Analyse & Kritik*, *26*, 279-301.

O'Sullivan, D. (2011). What is an encyclopedia? From Pliny to Wikipedia. In G. Lovink & N. Tkacz (Eds.), *Critical point of view: A Wikipedia reader* (pp. 34-49). Amsterdam, The Netherlands: Institute of Network Cultures.

Pasmore, W. A. (1988). *Designing effective organizations: The sociotechnical systems perspective*. New York, NY: Wiley.

Perez, J. C. (2007, February 17). Wikipedia breaks into U.S. top 10 sites. *PC World*. Retrieved from http://www.pcworld.com/article/129135/wikipedia_breaks_into_us_top_10_sites.html

Pfeifer, R., Bongard, J., & Grand, S. (2007). *How the body shapes the way we think: A new view of intelligence*. Cambridge, MA: MIT Press.

Piaget, J. (1953). *The origins of intelligence in children*. London, England: Routledge and Kegan Paul.

Piaget, J. (1970). Piaget's theory. In P. Mussen (Ed.), *Carmichael's handbook of child psychology* (pp. 703-732). New York, NY: Wiley.

Pinch, T. J. & Bijker, W. E. (1987). The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other. In W. E. Bijker, T. P. Hughes, & T. Pinch (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology* (pp. 17-50). Cambridge, MA: MIT Press.

Plesner, U. (2009). An actor-network perspective on changing work practices: Communication technologies as actants in newswork. *Journalism, 10*(5), 604-626.

Poe, M. (2006, September). The hive. *The Atlantic*. Retrieved from http://www.theatlantic.com/magazine/archive/2006/09/the-hive/5118/5/

Por, G. (2008). Interview with Professor Pierre Levy. In M. Tovey (Ed.), *Collective intelligence: Creating a prosperous world at peace* (pp. 283-292). Oakton, VA: Earth Intelligence Network.

Preda, A. (2006). Socio-technical agency in financial markets: The case of the stock ticker. *Social Studies of Science, 36*(5), 753-782.

Putnam, R. D. (1995). Bowling alone: America's Declining Social Capital. *Journal of Democracy, 6*(1), 65-78.

Rao, L. (2011, August 19). WordPress now powers 22 percent of new active websites in the U.S. *TechCrunch*. Retrieved from http://techcrunch.com/2011/08/19/wordpress-now-powers-22-percent-of-new-active-websites-in-the-us/

Reagle, J. M. (2010). *Good faith collaboration: The culture of Wikipedia*. Cambridge, MA: MIT Press.

Redden, J. (2008, March 4). Startup mines for riches in collaboration software. *Portland Tribune*. Retrieved from http://www.portlandtribune.com/rethinking /story.php?story_id=120430910578805900

Rheingold, H. (1993). *The virtual community: Homesteading on the electronic frontier*. Reading, MA: Addison-Wesley.

Rheingold, H. (2000). *The virtual community: Homesteading on the electronic frontier*, (rev ed.). Retrieved from http://www.caracci.net/dispense_enna/ The%20Virtual%20Community%20by%20Howard%20Rheingold_%20Table %20of%20Contents.pdf

Ricardo, D. (1817). *The principles of political economy & taxation*. London, England: J.M. Dent & sons, ltd.

Robu, V., Halpin, H., & Shepherd, H. (2009). Emergence of consensus and shared vocabularies in collaborative tagging systems. *ACM Transactions on the Web, 3*(4). Retrieved from http://portal.acm.org/citation.cfm?id=1594173.1594176

Rogers, R. (2009). *The end of the virtual*. Amsterdam, The Netherlands: Vossiuspers Uva.

Rojas, R., & Hashagen, U. (2000). *The first computers: History and architectures*. Cambridge, MA: MIT Press.

Ropohl, G. (1999). Philosophy of socio-technical systems. *Society for Philosophy and Teaching, 4*(3), 59-70.

Rosen, R. J. (2011, September 3). So, was Facebook responsible for the Arab Spring after all? *The Atlantic*. Retrieved from http://www.theatlantic.com/technology/archive/ 2011/09/so-was-facebook-responsible-for-the-arab-spring-after-all/244314/

Rothenberg, R. B. (1995). Commentary: Sampling in social networks. *Connections, 18*(1), 104-110.

Royal, C., & Kapila, D. (2008). What’s on Wikipedia, and What's Not . . . ?: Assessing Completeness of Information. *Social Science Computer Review, 27*(1), 138-148.

Rubin, H. J. & Rubin, I. S. (2005). *Qualitative interviewing: The art of hearing data*. Thousand Oaks, CA: Sage.

Salminen, J. (2012). Collective intelligence in humans: A literature review. Proceedings from *Collective Intelligence 2012*. Cambridge, MA.

Sanger, L. (2001a, January 10). Let's make a wiki. Retrieved from http://web.archive.org/web/20030822044513/http://www.nupedia.com/pipermail/nupedia-l/2001-January/000676.html

Sanger, L. (2001b, July 25). Britannica or Nupedia? The future of free encyclopedias. Retrieved from http://www.kuro5hin.org/story/2001/7/25/103136/121/

Sanger, L. (2002, February). Announcement about my involvement in Wikipedia and Nupedia. Retrieved from http://meta.wikimedia.org/wiki/Announcement_about_my_involvement_in_Wikipedia_and_Nupedia--Larry_Sanger

Sawyer, S., & Tapia, A. (2007). From findings to theories: Institutionalizing social informatics. *The Information Society, 23*(4), 263-275.

Schanze, J. (Director). (2010). *Plug and pray* [Motion picture]. Germany: Mascha Film.

Schuster, A. J. (2007). Intelligent computing everywhere. In A. J. Schuster (Ed.), *Intelligent computing everywhere* (pp. 3-24). London, England: Springer.

Schiller, Herbert. (1989). *Culture, Inc.: The corporate takeover of public expression*. NewYork, NY: Oxford University Press.

Scott, J. (1991). *Social network analysis*. London, England: Sage.

Segev, E. (2010). *Google and the digital divide: The bias of online knowledge*. Oxford, England: Chandos Publishing.

Shachaf, P., & Hara, N. (2010). Beyond vandalism: Wikipedia trolls. *Journal of Information Science*, *36*(3), 357 -370.

Shannon, C. E. (1948). A mathematical theory of communication. Bell Laboratories. Retrieved from http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf

Shannon, C. E. (1949). Communication theory of secrecy systems. *Bell System Technical Journal*, *28*(4), 656-715.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.

Shell, T. (2001, January 27). A few comments about Wikipedia. Retrieved from http://blog.gmane.org/gmane.science.linguistics.wikipedia.misc/month=20010101

Shirky, C. (2008). *Here comes everybody: The power of organizing without organizations*. New York, NY: Penguin Press.

Shirky, C. (2008, May 7). Gin, television, and social surplus. Retrieved from http://www.worldchanging.com/archives/008009.html

Signpost. (2011, July 25). News in brief: Advertising. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2011-07-25/News_and_notes

Silicon Valley Insider (2009, October 28). The SAI 50+: World's most valuable Internet startups. Retrieved from http://www.businessinsider.com/sai-50-2009

Sipress, A. (2007, March 5). Open call from the Patent Office. *The Washington Post*. Retrieved from http://www.washingtonpost.com/wp-dyn/content/article/2007/03/04/AR2007030401263.html

Sismondo, S. (2004). *An introduction to science and technology studies*. Malden, MA: Blackwell Pub.

Slater, R. (1987). *Portraits in silicon*. Cambridge, MA: MIT Press.

Smith, A. (1759). *The theory of moral sentiments*. London, England: Printed for A. Millar.

Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations*. London, England: W. Strahan and T. Cadell.

Smythe, D. (1977). Communications: Blindspot of western Marxism. *Canadian Journal of Political and Social Theory, 1*(3), 1-20.

Snow, M. (2005a, January 17). New servers to deal with performance issues. *The Wikipedia Signpost*. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2005-01-17/Server_order

Snow, M. (2005b, September 26). Wikimedia adds second full-time employee. *The Wikipedia Signpost*. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2005-09-26/Wikimedia_hire

Snow, M. (2007). Wikipedia officially a top ten site, according to Alexa. *The Wikipedia Signpost*. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2007-04-02/News_and_notes

Soderberg, J. (2002). Copyleft vs. copyright: A Marxist critique. *First Monday, 7*(3). Retrieved from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/rt/printerFriendly/938/860

Sollisch, J. (2012, March 22). On the death of Encyclopaedia Britannica: All authoritarian regimes eventually fall. *Christian Science Monitor*. Retrieved from http://www.csmonitor.com/Commentary/Opinion/2012/0322/On-the-death-of-Encyclopaedia-Britannica-All-authoritarian-regimes-eventually-fall

squid-cache.org. (2012). Squid: Optimising Web delivery. Retrieved from http://www.squid-cache.org/

Spivack, N. (2008). Harnessing the collective intelligence of the World-Wide Web. In M. Tovey (Ed.), *Collective intelligence: Creating a prosperous world at peace* (pp. 293-304). Oakton, VA: Earth Intelligence Network.

Spearman, C. (1927). *The nature of intelligence and the principles of cognition*. London, England: MacMillan.

Starosielski, N. (2011). Underwater flow. *FLOW, 15*(1). Retrieved from http://flowtv.org/2011/10/underwaterflow/

Steeves, H. L. & Wasko, J. (2002). Feminist theories and political economy: Toward a friendly alliance. In E. R. Meehan & E. Riordan (Eds.), *Sex & money: Feminism and political economy in the media* (pp. 16-29). Minneapolis, MN: University of Minnesota Press.

Stvilia, B., Mon, L., Yi, Y. (2009). A model for online consumer health information quality. *JASIST*, *60*(9), 1781-1791.

Sunstein, C. R. (2006). *Infotopia: How many minds produce knowledge*. Oxford, England: Oxford University Press.

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. New York, NY: Doubleday.

Sutter, J. D. (2009, August 26). Wikipedia: No longer the Wild West. *CNN*. Retrieved from http://edition.cnn.com/2009/TECH/08/26/wikipedia.editors/

Tankard, J. W., & Royal, C. L. (2005). What's on the web—And what's not. *Social Science Computer Review, 23*(3), 360-370.

Tapscott, D., & Williams, A. D. (2006). *Wikinomics: How mass collaboration changes everything*. New York, NY: Portfolio.

Technology Review. (2012). On topic: Arab Spring. Retrieved from http://www.technologyreview.com/ontopic/arabspring/

Terdiman, D. (2004, October 1). Solution for the Slashdot Effect? *Wired*. Retrieved from http://www.wired.com/science/discoveries/news/2004/10/65165

Terranova, T. (2004). *Network culture: Politics for the information age*. London, England: Pluto Press.

Tkacz, N. (2010). Wikipedia and the politics of mass collaboration. *PLATFORM: Journal of Media and Communication, 2*(2), 40-53.

Tian, Y. J. (2009). *Re-thinking intellectual property: The political economy of copyright protection in the digital era*. London, England: Routledge-Cavendish.

Tovey, M. (2008). *Collective intelligence: Creating a prosperous world at peace*. Oakton, VA: Earth Intelligence Network.

Travers, J. & Milgram, S. (1969). An experimental study in the small world problem. *Sociometry, 35*(4), 425-443.

Trist, E. L. (1959). On socio-technical systems. In W. A. Pasmore & J. J. Sherwood (Eds.), *Sociotechnical systems: A sourcebook* (pp. 43-57). San Diego, CA: University Associates, Inc.

Trist, E., & Bamforth, K. (February 01, 1951). Some Social and Psychological Consequences of the Longwall Method of Coal-Getting. *Human Relations,* 4(1), 3-38.

Turing, A.M. (1936). On computable numbers, with an application to the Entscheidungs problem". *Proceedings of the London Mathematical Society, 2*(42), 230–65.

Turing, A. M. (1950, October 1). Computing Machinery and Intelligence. *Mind*, *59*(236): 433-460.

United Nations University. (2010, March). Wikipedia survey—Overview of results. Retrieved from http://www.wikipediasurvey.org/docs/ Wikipedia_Overview_15March2010-FINAL.pdf

United Press International. (2008, March 12). Wikipedia struggles with funding. Retrieved from http://www.upi.com/Top_News/2008/03/12/Wikipedia-struggles-with-funding/UPI-41721205349952/

Utz, S. (2009). “Egoboo” vs. altruism: the role of reputation in online consumer communities. *New Media & Society*, *11*(3), 357 -374.

Vaidhyanathan, S. (2011). *The Googlization of everything: (and why we should worry)*. Berkeley, CA: University of California Press.

Van Dijck, J., & Nieborg, D. (2009). Wikinomics and its discontents: A Critical analysis of Web 2.0 business manifestos. *New Media & Society, 11*(5), 855-874.

Vego-Redondo, F. (2007). *Complex social networks*. Cambridge, England: Cambridge University Press.

Venturini, T. (2010). Diving in magma: How to explore controversies with actor-network theory. *Public Understanding of Science, 19*(3), 258-273.

Vibber, B. (2006, April 28). Wikipedia and MediaWiki. Retreived from http://www.youtube.com/watch?v=DWef69ItVrU

Viégas, F. B., Wattenberg, M., & Dave, K. (2004). Studying cooperation and conflict between authors with history flow visualizations. Proceedings from *Computer-human Interaction 2004*. Vienna, Austria.

Viégas, F. B., Wattenberg, M., & McKeon, M. M. (2007). The hidden order of Wikipedia. *Proceedings of the 2nd international conference on Online communities and social computing* (pp. 445–454).

Von Neumann, J. (1945). First draft of a report on the EDVAC. Retrieved from http://www.youtube.com/watch?v=DWef69ItVrU

Wales, J. (2001, October 27). Statement of principles. Retrieved from http://nostalgia.wikipedia.org/w/index.php?title=Jimbo_Wales/Statement_of_principles&oldid=75340

Wales, J. (2003). Announcing Wikimedia Foundation. Retrieved from http://lists.wikimedia.org/pipermail/wikipedia-l/2003-June/010743.html

Wales, J. (2008, March 7). Wikipedia:Advertisements. Retrieved from http://en.wikipedia.org/wiki/User_talk:Jimbo_Wales/Archive_35#Wikipedia:Advertisements

Walsh, C. & Apperley, T. (Eds.). (2011). Interview with Nathaniel Tkacz on behalf of the Critical Point of View collective. Retrieved from http://www.digitalcultureandeducation.com/uncategorized/dce_interview_nathaniel_tkacz_2011_html/

Wasko, J. (2003). *How Hollywood works*. London, England: Sage.

Wasko, J. (2004). The political economy of communications. In J. Downing. (Ed.), *The SAGE handbook of media studies* (pp. 309-329). Thousand Oaks, CA: Sage.

Wasko, J. & Erickson, M. P. (2009). The political economy of YouTube. In P. Snickars & P. Vonderau (Eds.), *The YouTube reader* (pp. 372-386). Stockholm, SW: National Library of Sweden.

Wasko, M. M., & Faraj, S. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly, 29*(1), 35-57.

Wasserman, S., Scott, J., & Carrington, P. J. (2005). Introduction. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 1-7). Cambridge, England: Cambridge University Press.

Watts, D. J. (1999). Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology, 105*(2), 493-527.

Watts, D. J. (2003). *Six degrees: The science of a connected age*. New York, NY: Norton.

Weiss, A. S., & Domingo, D. (2010). Innovation processes in online newsrooms as actor-networks and communities of practice. *New Media & Society, 12*(7), 1156-1171.

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. San Francisco, CA: W.H. Freeman.

Wiener, N. (1961). *Cybernetics; or, Control and communication in the animal and the machine*. New York, NY: MIT Press.

Wikimedia Commons. (2011, January 1). WalesCalltoAction.ogv. Retrieved from http://ten.wikipedia.org/wiki/File:WalesCalltoAction.ogv

Wikimedia Foundation. (2005). Press releases/Kennisnet supports Wikimedia projects. Retrieved from http://wikimediafoundation.org/wiki/Press_releases/Kennisnet_supports_Wikimedia_Projects

Wikimedia Foundation. (2006). Resolution:Advisory board. Retrieved from http://wikimediafoundation.org/wiki/Resolution:Advisory_board

Wikimedia Foundation. (2007). Resolution:Appointment of Sue Gardner as ED. Retrieved from http://wikimediafoundation.org/wiki/Resolution:Appointment_of_Sue_Gardner_as_ED

Wikimedia Foundation. (2008). 2007-2008 Annual Report. Retrieved from http://upload.wikimedia.org/wikipedia/foundation/2/26/WMF_20072008_Annual_report._high_resolution.pdf

Wikimedia Foundation. (2009). 2009-2009 Annual Report. Retrieved from http://upload.wikimedia.org/wikipedia/foundation/a/a4/WMF_Annual_Report_20082009_online.pdf

Wikimedia Foundation. (2010a). 2009-2010 Annual Report. Retrieved from http://upload.wikimedia.org/wikipedia/commons/9/9f/AR_web_all-spreads_24mar11_72_FINAL.pdf

Wikimedia Foundation. (2010b). Mission statement. Retrieved from http://wikimediafoundation.org/wiki/Mission_statement

Wikimedia Foundation. (2011a). 2010-2011 Annual Report. Retrieved from http://upload.wikimedia.org/wikipedia/commons/4/48/WMF_AR11_SHIP_spreads_15dec11_72dpi.pdf

Wikimedia Foundation. (2011b). 2010-2011 FAQ. Retrieve from http://upload.wikimedia.org/wikipedia/foundation/7/72/Audit_FAQ_2011_Final.pdf

Wikimedia Foundation. (2011c). From one of the creators of Wikipedia. Retrieved from http://wikimediafoundation.org/wiki/L11_KA6banner_Julytest_ST/en/US

Wikimedia Foundation. (2011d). Press releases/Stanton Foundation awards Wikimedia $3.6 million for technology improvements. Retrieved from https://wikimediafoundation.org/wiki/Press_releases/Stanton_Foundation_Awards_Wikimedia_$3.6_Million_for_Technology_Improvements

Wikimedia Foundation. (2011e). Wikimedia strategic plan: A collaborative vision for the movement through 2015. Retrieved from http://upload.wikimedia.org/wikipedia/commons/c/c0/WMF_StrategicPlan2011_spreads.pdf

Wikimedia Foundation. (2012a). Frequently asked questions. Retrieved from http://wikimediafoundation.org/wiki/FAQ/en

Wikimedia Foundation. (2012b). Staff and contractors. Retrieved from https://wikimediafoundation.org/wiki/Staff_and_contractors

Wikimedia Meta-Wiki. (2004). Board elections/2004/Results. Retrieved from http://meta.wikimedia.org/wiki/Board_elections/2004/Results

Wikimedia Meta-Wiki. (2011a). Fundraising 2011. Retrieved from http://meta.wikimedia.org/wiki/Fundraising_2011

Wikimedia Meta-Wiki. (2011b). Notes on good practices on Wikipedia research. Retrieved from http://meta.wikimedia.org/wiki/ Notes_on_good_practices_on_Wikipedia_research

Wikimedia Meta-Wiki. (2011c). Research:Understanding the editor/bot relationship. Retrieved from http://meta.wikimedia.org/wiki/ Research:Understanding_the_Editor/Bot_Relationship

Wikimedia Meta-Wiki. (2011d). Research:Wikipedia editors survey 2011. Retrieved from http://meta.wikimedia.org/wiki/Wikipedia_Editors_Survey_2011

Wikimedia Meta-Wiki. (2012a). Bot policy. Retrieved from http://meta.wikimedia.org/wiki/Bot_policy

Wikimedia Meta-Wiki. (2012b). Help:Unified login. Retrieved from http://meta.wikimedia.org/wiki/Help:Unified_login

Wikimedia Meta-Wiki. (2012c). Meta-Wiki. Retrieved from https://meta.wikimedia.org/wiki/Main_Page

Wikimedia Meta-Wiki. (2012d). Research Committee. Retrieved from http://meta.wikimedia.org/wiki/Research:Committee

Wikimedia Meta-Wiki. (2012e). Stewards. Retrieved from http://meta.wikimedia.org/wiki/Steward

Wikimedia Meta-Wiki. (2012f). Toolserver. Retrieved from http://meta.wikimedia.org/wiki/Toolserver

Wikimedia Meta-Wiki. (2012g). User groups. Retrieved from http://meta.wikimedia.org/wiki/User_groups

Wikimedia Meta-Wiki. (2012h). User:Sue Gardner. Retrieved from https://meta.wikimedia.org/wiki/User:Sue_Gardner

Wikipedia. (2005). Wikipedia:Wikipedia Signpost/2005-03-14/Page moves. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2005-03-14/Page_moves

Wikipedia. (2007a). Esperanza organization disbanded after deletion discussion. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2007-01-02/Experanza

Wikipedia. (2007b). User:CatherineMunro/Why am I here? Retrieved from http://en.wikipedia.org/wiki/User:CatherineMunro/Why_am_I_here%3F

Wikipedia. (2009a). User:Clifford Adams. Retrieved from http://en.wikipedia.org/wiki/User:Clifford_Adams

Wikipedia. (2009b). Wikipedia:There are no rules. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:There_are_no_rules

Wikipedia. (2010). Wikipedia:Size of Wikipedia. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia

Wikipedia. (2011a). User:Lee Daniel Crocker. Retrieved from http://en.wikipedia.org/wiki/User:Lee_Daniel_Crocker

Wikipedia. (2011b). Wikipedia:CamelCase and Wikipedia. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:CamelCase_and_Wikipedia

Wikipedia. (2012a). Collective intelligence. Retrieved from

Wikipedia. (2012b). Five pillars. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Five_pillars

Wikipedia. (2012c). Folksonomy. Retrieved from http://en.wikipedia.org/wiki/Folksonomy

Wikipedia. (2012d). History of Wikipedia. Retrieved from http://en.wikipedia.org/wiki/History_of_Wikipedia

Wikipedia. (2012e). LAMP (software bundle). Retrieved from http://en.wikipedia.org/wiki/LAMP_%28software_bundle%29

Wikipedia. (2012f). MediaWiki. Retrieved from http://en.wikipedia.org/wiki/MediaWiki

Wikipedia. (2012g). Nupedia. Retrieved from http://en.wikipedia.org/wiki/Nupedia

Wikipedia. (2012h). Pierre Larousse. Retrieved from http://en.wikipedia.org/wiki/Pierre_Larousse

Wikipedia. (2012i). Special:Statistics. Retrieved from http://en.wikipedia.org/wiki/Special:Statistics

Wikipedia. (2012j). User:ClueBot NG. Retrieved from http://en.wikipedia.org/wiki/User:ClueBot_NG

Wikipedia. (2012k). User:Dcoetzee/Willy on wheels: A case study. Retrieved from http://en.wikipedia.org/wiki/User:Dcoetzee/Willy_on_Wheels:A_Case_Study

Wikipedia. (2012l). User:Jimbo Wales/Statement of principles. Retrieved from http://en.wikipedia.org/wiki/User:Jimbo_Wales/Statement_of_principles

Wikipedia. (2012m). User:Snowolf/Adminbots. Retrieved from http://en.wikipedia.org/wiki/User:Snowolf/Adminbots

Wikipedia. (2012n). Wikipedia:Administrators. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Administrators

Wikipedia. (2012o). Wikipedia:Advertisements. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Advertisements

Wikipedia. (2012p). Wikipedia:AutoWikiBrowser. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:AutoWikiBrowser

Wikipedia. (2012q). Wikipedia:Barnstars. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Barnstars

Wikipedia. (2012r). WikipediaBiographies of living persons. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Biographies_of_living_persons

Wikipedia. (2012s). Wikipedia:Bot policy. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Bot_policy

Wikipedia. (2012t). Wikipedia:Bots. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Bots

Wikipedia. (2012u). Wikipedia:Bureaucrats. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Bureaucrats

Wikipedia. (2012v). Wikipedia:Edit count. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Edit_count

Wikipedia. (2012w). Wikipedia:Esperanza. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Esperanza

Wikipedia. (2012x). Wikipedia:FAQ/Technical. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:FAQ/Technical#History_of_Wikipedia_Hardware

Wikipedia. (2012y). Wikipedia: Featured articles. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Featured_articles

Wikipedia. (2012z). Wikipedia:Ignore all rules. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Ignore_all_rules

Wikipedia. (2012aa). Wikipedia:Namespace. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Namespace

Wikipedia. (2012bb). Wikipedia:No original research. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:No_original_research

Wikipedia. (2012cc). Wikipedia:Red link. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Red_link

Wikipedia. (2012dd). Wikipedia:Systemic bias. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Systemic_bias#cite_note-0

Wikipedia. (2012ee). Wikipedia talk:Bots/Archive. Retrieved from http://en.wikipedia.org/wiki/Wikipedia_talk:Bot_policy

Wikipedia. (2012ff). Wikipedia:User access levels. Retrieved from http://en.wikipedia.org/wiki/Help:Bureaucrat_rights#cite_note-0

Wikipedia. (2012gg). Wikipedia:Why create an account? Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Why_create_an_account%3F

Wikipedia. (2012hh). Wikipedia:Wikipedians. Retrieved from http://en.wikipedia.org/wiki/Wikipedia:Wikipedians

Wilkinson, D., & Huberman, B. (2007). Cooperation and quality in Wikipedia. *WikiSym'07*, Montreal, Canada.

Williams, C. (2011, January 28). How Egypt shut down the Internet. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/news/worldnews/africaandindianocean/egypt/8288163/How-Egypt-shut-down-the-internet.html

Willy on Wheels. (2005, October 30). Clearing my conscience. Retrieved from http://lists.wikimedia.org/pipermail/wikien-l/2005-October/031780.html

Wortham, J. (2012, January 16). Wikipedia to go dark on Wednesday to protest bills on Web piracy. *The New York Times*. Retrieved from http://bits.blogs.nytimes.com/2012/01/16/wikipedia-plans-to-go-dark-on-wednesday-to-protest-sopa/

Wu, T. (2010). *The master switch: The rise and fall of information empires*. New York, NY: Alfred A. Knopf.

Zachte, E. (2011). Wikipedia statistics: Bot activity. Retrieved from http://stats.wikimedia.org/EN/BotActivityMatrix.htm

Zachte, E. (2012a). Wikimedia report card February 2012. Retrieved from http://stats.wikimedia.org/reportcard/#fragment-32

Zachte, E. (2012b). Wikipedia Statistics. Retrieved from http://stats.wikimedia.org/EN/TablesWikipediansEditsGt100.htm

Zittrain, J. (2008). *The future of the Internet and how to stop it*. New Haven, CT: Yale University Press.