

Coordination in Commons-Based Peer Production Communities:  
Evaluation, Analysis, and Design for Emergent Systems

Michael D. Gilbert

A dissertation  
submitted in partial fulfillment of the  
requirements for the degree of

Doctor of Philosophy

University of Washington  
2015

Reading Committee:  
Mark Zachry, Chair  
David W. McDonald  
Benjamin Mako Hill  
Kate Starbird

Program Authorized to Offer Degree:  
Human Centered Design & Engineering

©Copyright 2015  
Michael D. Gilbert

University of Washington

**Abstract**

Coordination in Commons-Based Peer Production Communities:  
Evaluation, Analysis, and Design for Emergent Systems

Michael D. Gilbert

Chair of the Supervisory Committee

Professor Mark Zachry

Human Centered Design & Engineering

Traditional organizations are often relatively structured, involving hierarchies of individual agents driving progress through production, with well-defined boundaries between internal requirements and external expectations and with explicit goals set to drive work processes to their ideal end. Modern commons-based peer production communities, however, exist in a different reality, both conceptually and physically. These communities of collaborators are often identified by their distributed agents acting among distributed teams, functioning without formal hierarchy, each contributing to disparate projects with divergent goals, each with myriad individual motivations, all leading to an emergent order that arises not from managerial decree but from the mass of actors participating in the melee. In these communities, coordination is no longer a top-down process that can be centrally controlled. It is instead a product of emergent order; it is a product of the ongoing and ever changing narrative of the individual actors contributing to and ultimately shaping those communities they interact within. This dissertation examines such coordination within one of the most visible commons-based peer production communities – Wikipedia. Within Wikipedia, millions of volunteers have contributed to create one of the largest and most visited resources in the world, a free online encyclopedia with the goal to make available the sum of all human knowledge. While prior research has contributed to theoretical understanding of coordination in organizations, no such theory exists for modern commons-based peer production communities. This dissertation presents such a theory, grounded in observations of the agents within that community, with the capacity to explicate the many discrete interactions that inform the whole, to identify the means by which those interactions succeed or fail, and to ultimately provide a theoretical foundation for future design interventions. Finally, to validate this theory I present the Virtual Team Explorer, a theory-driven design study resulting in a prototype application intended to facilitate the coordination of distributed actors within WikiProjects in Wikipedia, along with the tools and requisite documentation that make such an intervention possible.

[This page intentionally left blank]

A dedication:

For the patient. And the curious.

# Acknowledgements

A heartfelt thanks is deserved, and one readily extended, for all the individuals and actors who engaged with and supported this work. First, I would like to thank my advisor, Professor Mark Zachry, for his consistent and tireless support throughout my time in the program, for his encouragement of both the ideas that led to the work in this dissertation as well as the ideas that still ruminate, awaiting their time to arise. As well, I would like to thank Professor David McDonald, for the opportunities he provided, for challenging good ideas and improving bad ones, and for the patience he exhibited in the face of unending streams of emails and questions from me. I would also like to thank the members of my dissertation committee, among them Mark and David: Kate Starbird and Benjamin Mako Hill. From early on these individuals provided a basis for exploration and inquisition, each lending their efforts to support the provenance and development of the research which follows, and to the researcher striving to create it.

Equally deserving of credit are those that have endured my run-on sentences and dangling participles, as well my enthusiasm which occasionally masks my naiveté. Among them, much credit goes to Jonathan Morgan, my academic brother, for his insight and encouragement, and Elizabeth Churchill, for consistently driving me to become a better researcher and never failing to believe that drive was justified. I would also like to thank the members of my research group, without whom this work would not have been possible: Jessica Bao, Margaret Lyons, Greg Robison, and Ruiyi Zhou. I would also like to thank my HCDE brethren for their support, encouragement, and occasional happy hour, among them Taylor Scott, Daniel Perry, and Doug Divine. I would like to thank the National Science Foundation (Grants IIS-0811210 and IIS-1162114) for funding this research, and the larger HCDE and University of Washington for providing me with the opportunity and encouragement to complete that research.

Finally, I thank my family, specifically brother, Scott Gilbert, and my mother and father, Barbara Gilbert and Michael Gilbert, for their encouragement, support, and their enduring belief that pretty much most of the time, pretty much anything is possible. Lastly, I thank my friends, Joe and Jen Vollan, Nichole Rathburn, and Corey McDonough, for listening to and participating in the many late night conversations regarding arcane subjects, both related to this work and not.

# Table of Contents

Abstract .....	3
Acknowledgements .....	6
List of Tables.....	10
List of Figures .....	11
Introduction .....	13
Communities of Collaboration: Situating Modern Conceptions of Community in Socio-Technical Systems .....	14
Peer Production of the Sum of all Human Knowledge: A Brief Introduction to the Curious Case of Wikipedia.....	17
Structures of Collaboration: Projects, Problems, and Solutions .....	19
Structure of this Dissertation & Research Goals.....	20
A Brief Note on Authorial Voice .....	23
<b>Part I – Overview of Prior Research: Foundations for Development.....</b>	<b>24</b>
1.1 Understanding Wikipedia.....	24
1.1.1 Understanding Wikipedia: How Measurement Shapes Participation and Value .....	25
1.1.2 Prior Conceptions of Coordination in Wikipedia .....	27
1.2 Collaboration in Commons-Based Peer Production communities .....	33
1.3 Coordination Theory – a Brief Primer .....	35
1.4 Conclusion to Part I.....	36
<b>Part II – Grounding Coordination: Understanding Interactions in Context.....</b>	<b>38</b>
2.1 Study 1: Managing complexity: Strategies for Group Awareness and Coordinated Action in Wikipedia.....	38
2.1.1 Introduction – Exploring Opportunities for Ambient Information Transfer.....	39
2.1.2 Two Key Awareness Mechanisms in WikiProjects.....	39
2.1.3 Research Hypotheses: Assessing the Impact of the Hot Articles Situational Awareness Mechanism in WikiProjects.....	41
2.1.4 Research Method: Data Collection and Analysis for the Hot Articles Tool.....	41
2.1.5 Results: Analyzing the Impact of Awareness Tools on Levels of Contribution.....	45
2.1.6 Discussion .....	48
2.2 Conclusion to Study 1 .....	51
2.3 Study 2: Supporting Tool-Mediated Coordination of Virtual Teams in Complex Systems.....	51
2.3.1 Introduction: Situating Tool-Mediation in Current Practice.....	52
2.3.2 Related Work: Tool-mediation in commons-based peer production communities.....	53
2.3.3 Data Collection & Analysis.....	55

2.3.4 Discussion .....	60
2.4 Conclusion to Study 2 .....	62
2.5 Part II Conclusion .....	63
<b>Part III - “Opportunistic Ad-Hocracies”: Towards a Theory of Coordination for Commons-Based Peer Production Communities.....</b>	<b>64</b>
3.1 Introduction: Refocusing the Problem .....	64
3.2 Research Questions .....	65
3.3 Research Method.....	65
3.3.1 A Directed Content Analysis of Coordination.....	66
3.3.2 Grounded Theory in Practice.....	68
3.4 Evaluating the Theory of Coordination for Commons-Based Peer Production Communities .....	77
3.4.1 Quantitative Evaluation of Coordination Framework.....	78
3.4.2 Evaluation of Coordination Framework: Dependency Trajectory Maps of Interview Transcript Scenarios .....	78
3.4.3 Evaluation of Coordination Framework: Dependency Trajectory Maps of WikiProject Interactions .....	82
3.5 Future Work .....	87
3.5.1 Justification for this Theory of Coordination for Commons-Based Peer Production Communities (or, why we need more theories of coordination).....	88
3.6 Discussion .....	89
3.7 Conclusion to Part III.....	89
<b>Part IV – Creation of the Virtual Team Explorer: Theory applied.....</b>	<b>91</b>
4.1 Introduction for the Intervention .....	92
4.2 Related Work: Consolidating Interpretations of Coordination.....	92
4.2.1 Situating the Virtual Team Explorer: Review and Recap.....	93
4.2.2 Reviewing Prior Works through the Lens of a Theory of Coordination for Commons-Based Peer Production Communities .....	96
4.3 System Architecture .....	98
4.3.1 Data Structures for Cooperative Design .....	99
4.3.2 Platforms for Deployment .....	100
4.3.3 Extensibility of Approach.....	101
4.3.4 Synchronous-Pedia .....	103
4.4 System Design.....	103
4.4.1 Supporting Collaborative Work.....	103
4.4.2 Methodology: Situating the Virtual Team Explorer within a Theory of Coordination.....	110
4.5 Evaluating the Virtual Team Explorer .....	111
4.6 Designing Beyond MediaWiki.....	115

4.7 Next Steps .....	116
4.8 Conclusion to Part IV.....	117
<b>Part V: Conclusion.....</b>	<b>119</b>
5.1 Future Work.....	123
References.....	126
About the Author.....	132
Appendices .....	133
Appendix A: Coordination Codesheet Description for Directed Content Analysis .....	133
Appendix B: Coordination Codesheet Description from Grounded Theory Analysis .....	140
Appendix C: Interview protocol.....	144
Appendix D: Example interview.....	146
Appendix E: Testing Coder Agreement of our Theory of Coordination for Commons-Based Peer Production Communities with Scenarios .....	154
Appendix F: Instructions for Task Decomposition and Dependency Mapping to Create Concept Maps from Scenarios Derived from Interview Transcripts.....	155
Appendix G: Virtual Team Explorer – WikiProject Scenario Survey .....	160
Appendix H: Identifying Automated Tool and Bot Edits .....	170
Appendix I: Towards a Class-Based Model of Information Organization in Wikipedia: Data Structures Utilizing Semi-Structured Text.....	171

# List of Tables

Table 1. Malone and Crowston’s common dependencies and example alternative coordination mechanisms for dealing with them. Originally from (Malone & Crowston, 1994) .....	36
Table 2. Non-bot edits to WikiProject pages identified by user and revision comment. Not represented are 19 other tools .....	57
Table 3. Non-bot edits to WikiProject pages identified by concurrent edits which occurred $\leq 5$ seconds apart from one another .....	58
Table 4. Prevalence of coordination mechanisms utilized in automated and semi-automated tools in the project space. Percentages given in top-level categories are global, sub-category percentages are relative to each category. Categories observed in practice but not included in the original coordination framework are included under the existing framework .....	60
Table 5. Details of interview schedule .....	66
Table 6. General guidelines for evaluating the results of Cohen’s Kappa (Landis & Koch, 1977) .....	68
Table 7. Sample quotes identifying specific dependencies from interview transcripts. Descriptions under each quote are taken from memos written during the initial interview, transcription, or coding process (Charmaz, 2014) .....	72
Table 8. Coordination framework for analysis of coordination in commons-based peer production communities .....	76
Table 9. WikiProjects selected for operationalized evaluation of dependency trajectories. The Potential Dependencies Observed includes all dependencies identified over the entire life of the WikiProject. WikiProjects were selected from those encountered in the interview transcripts.....	84
Table 10. Overview of prior coordination studies .....	96
Table 11. Results of the survey evaluation of the Virtual Team Explorer prototype .....	112
Table 12. Comments given by survey participants for the Virtual Team Explorer prototype scenarios .....	112
Table 13. The most frequently transcluded Templates and Modules in the English Language Wikipedia..	175

# List of Figures

Figure 1. Timeline showing the multiple branches and individual operating systems (over 30 in all) that resulted from the release of the initial Linux kernel. Source: Wikimedia commons, Linux_Distro_Timeline .....	16
Figure 2. Total edits per week to the English Wikipedia across all namespaces by both registered (blue) and unregistered (orange) users.....	18
Figure 3. User (non-bot) edits per week to the English Wikipedia, by namespace.....	19
Figure 4. Bot edits per week to the English Wikipedia, by namespace.....	19
Figure 5. The first revision of the Wikipedia article “2013 Rosario gas explosion”.....	32
Figure 6. The Hot Articles page as it would appear on WikiProject pages that subscribe to the tool. Shown above are pages as ranked on WikiProject Feminism (left) and WikiProject Cats (right) .....	40
Figure 7. Gantt chart showing the projects included in the analysis, as well as the time span of the analysis. The black line in the center of each timeline represents the date the Hot Article bot was activated. The period before the bot was active (the “pseudo” period) was included to more effectively determine if changes in article editing patterns was due to the HA bot or other factors .....	42
Figure 8. Means box plots of total edits to Hot Article and Talk Page links and Unique editors to those links. The link period (pre or post) defines the period when project members made the edits – either in the seven days prior to (pre) or following (post) the time when the link was posted .....	45
Figure 9. Means box plots of ratio of total edits to unique editors for both edits to articles posted by the Hot Articles bot (left) and edits to articles posted to project talk pages (right). The edit type defines both the period when the edit was made (“pre” or “post”, either in the seven days prior to or following the link being added), and whether the edit was made by a project member (“mem” or “non”).....	48
Figure 10. Density plot of bot edits to WikiProjects (left) and non-WikiProjects (right). Red represents the Wikipedia namespace (4) and blue represents the Wikipedia Talk namespace (5). X-axis is square root of edit counts .....	56
Figure 11. Concept map of dependency trajectories derived from transcript scenarios. Dependencies identified for individual scenarios (left) along with the concept map resulting from combining each of those scenarios (right). The combined map is additionally showing betweenness and closeness scores for each dependency .....	81
Figure 12. Dependency trajectory maps from WikiProject Classroom coordination member interactions. Top-left and right indicate trajectories for individual users (two project members with the greatest number of observed potential dependencies), while the bottom image includes the combined observed dependencies for the entire WikiProject .....	85
Figure 13. Dependency trajectory maps from WikiProject Chemistry member interactions. Top-left and right indicate trajectories for individual users, while the bottom image includes the combined observed dependencies for the entire WikiProject .....	86
Figure 14. Recommendations for tasks to complete created by the SuggestBot tool. Articles shown are related to the Human Computer Interaction Articles category within Wikipedia.....	94
Figure 15. Example Task List from WikiProject Breakfast .....	99
Figure 16. Diagram from (Hill, 2013). While Everything2 and H2G2 are still active, our platform selection aimed to mirror Wikipedia on the Novelty vs. Familiarity spectrum. Reproduced with permission from the author.....	101
Figure 17. The project browser within the VTE provides an entry point for new users to explore active WikiProjects along a number of metrics .....	104

Figure 18. Force-directed, edge-bundled network graph showing both the explicit members (blue nodes) and procedural members (grey nodes) within WikiProject Chemistry. Red nodes indicate edits to articles with the scope of the project, yellow nodes indicate article edits outside the scope of the project (as defined by the category structure).....	107
Figure 19. Task browser in the VTE (top) and as displayed in the WikiProject page (bottom), both using the same WikiText to render. The ability to embed arbitrary metadata in task invocations means that a far greater amount of information (subtasks, for example) can be included without complicating the in-wiki representation .....	109
Figure 20. Example presentation of the article browser within the VTE, with the above showing the project and project talk page activity for WikiProject Biography, along with the top three most active articles by revision count for the last 30 days under the scope of the project.....	110
Figure 21. An example of the instantiation and presentation of the infobox template for the element Oxygen. The cell on the left contains the text used to define the template attributes, the image on the right is part of the resulting presentation .....	173
Figure 22. (a) Distribution of pages across namespaces, (b) Template and Module transclusions on all pages, split by destination namespace .....	175

# Introduction

Thoughts of *community* often conjure very specific feelings – be it late night barbecues with close friends, neighbors offering fresh basil from their sidewalk garden, groups of like-minded individuals gathering around tables to game, to commiserate, to share their life stories, or co-workers leaning casually over cubicle walls to discuss business, to discuss family, or to discuss lunch plans as a momentary respite from a long day. These communities and others similar are those invoked by popular media when recalling older days and simpler times, involving kids running down spotless streets towards playgrounds with carefully manicured lawns, and with family announcements delivered over the intricacies of copper wire in convoluted games of telephone tag. In these imaginings, information broadcasts are one-way, spoken by venerable sources endeared to these communities through trust and impartiality, giving proclamations of both victories and defeats in grave tones capable of calming the most unquiet of souls.

Things have of course, in many very important and dramatic ways, changed. Increasingly, the communities we spend the majority of our time in look nothing like these prior imaginings. News is no longer merely consumed, but interpreted, argued, and rebroadcast *ad infinitum* among social networks and downstream aggregators. With the emergence of increasingly capable technologies and information infrastructures has come the capacity for these communities to exist in a far more ethereal state – friendships can be forged and maintained spanning geographic boundaries based on nothing but passing interests, families can now instantaneously broadcast the latest pictures of toddlers toddling to the world in a format where even those with only a mild passing interest can express their amazement with a single click of a “Like” button, and the casual meetup *qua* friendly barbecue or block party can now be organized, evangelized, and coordinated entirely without its participants being in the same room.

In many ways, the technologies that have made this progression of our notion of “community” possible have done so through removing the boundaries necessary to enact meaningful communication. For instance, in the event of a community emergency it is no longer necessary to participate in elaborate phone trees, wherein each person is responsible for calling multiple others who, in turn, continue calling multiple others until the message is received by all. In a far less tedious process, now, any message can be broadcast to all other members of that community, each an individual node in a completely connected network over myriad potential channels, each with the capacity to receive and interact with that message in any manner befitting their individual motivations. To take a more concrete example, Monroy-Hernandez (2013) explored how modern microblogging platform Twitter has been used by average citizens to inform and caution one another about violent outbreaks occurring in certain towns, effectively co-opting the technology to become “curators,” or “war correspondents” with the capacity to reach countless numbers of people both within those communities as well as far beyond. Similar work by Starbird and Palen (2011) examined how individual volunteers, or “voluntweeters,” used the same Twitter platform to organize and coordinate rescue efforts in the wake of the 2010 Haiti earthquake. Other work by Skeels and Grudin (2009) examined how similar services – specifically, Facebook and LinkedIn – have been used within professional environments to both

communicate information related to those environments as well as to maintain social networks that span the chasm between the personal and the professional.

These examples are not given as a precursor to an examination on the myriad ways that these various platforms can be adopted and adapted for individual use. Nor are they a prelude to an argument to delineate the varied means by which distributed individuals collaborate, be it netWORKers in their Intensional Networks (Nardi, Whittacker, Schwarz, 2002), nor various communities of practice (Wenger, 2000) or varied Actor-Networks (Latour, 1996). These examples, rather, are given as a very simple means of distinguishing between how the notion of “community” may have once been *enacted* compared to the many ways in which it may be enacted now. In other words, bracketing for a moment any potential disagreement regarding *why* people interact within their respective “communities,” including their motivations, inclinations, or predispositions, these examples are meant to illuminate *how* those varied interactions may take place. More simply, these examples are given as a prelude to framing the shift as conceptions of community move from the geographical or interpersonal to the socio-technical. In this conception, the community exists both within and is partially comprised of the technologies that enable its interactions. The enactment and maintenance of these communities, then, are the result of the collective action of all those who interact within them, in addition to the technologies that shape those interactions, the quality control mechanisms that ensure their ongoing health, the design of experiences that impact the nature of participation, as well as the very means that information and interaction within those communities are recorded and displayed to its various members and the tools that buttress its existence. With this clarification, the next section will introduce the specific communities under examination within this dissertation, as well as provide the necessary background and language to more formally address and understand those communities and the interactions within them.

## **COMMUNITIES OF COLLABORATION: SITUATING MODERN CONCEPTIONS OF COMMUNITY IN SOCIO-TECHNICAL SYSTEMS**

In “Coase’s Penguin, or, Linux and the Nature of the Firm,” Yochai Benkler (2002) introduced the term commons-based peer production (CBPP) to refer to a new model of production, distinct from the *markets* and *firms* predating these new types of communities. While the structural characteristics that typify these new communities will be further expanded upon in Part I, it remains helpful to clarify the more general nature of first, what they are, second, what types of production occurs within them, and third, what are potential exemplars within the field.

Starting from the former, Benkler characterizes this new type of production through contradistinction with the prior mentioned markets and firms. Specifically, he states that both markets and firms exist to address a specific question: for any given individual agent in a collaborative production system at any given point in time, what should that agent be working on? In the case of markets, this question is answered through attaching specific prices to unique outcomes. In this instance, the individual worker has the capacity to identify tasks to complete given the reward (i.e., payment) for those tasks. In the case of firms, task assignment is completed through managerial decree in that individual agents are responsible for completing the work assigned to them, whereby each assignment is given a unique weight to facilitate the completion of a desired outcome. For both markets and firms, there is a transaction

cost associated with that process of production. For markets, the transaction cost is the price associated with identifying and producing a desired outcome. For firms, it is the cost of organizing that production through an internal structure. Markets, Benkler argues, are selected “when the gains from doing so, net of transaction costs, exceed the gains from doing the same thing in a managed firm, net of organization costs.” When those criteria are not met, firms are selected.

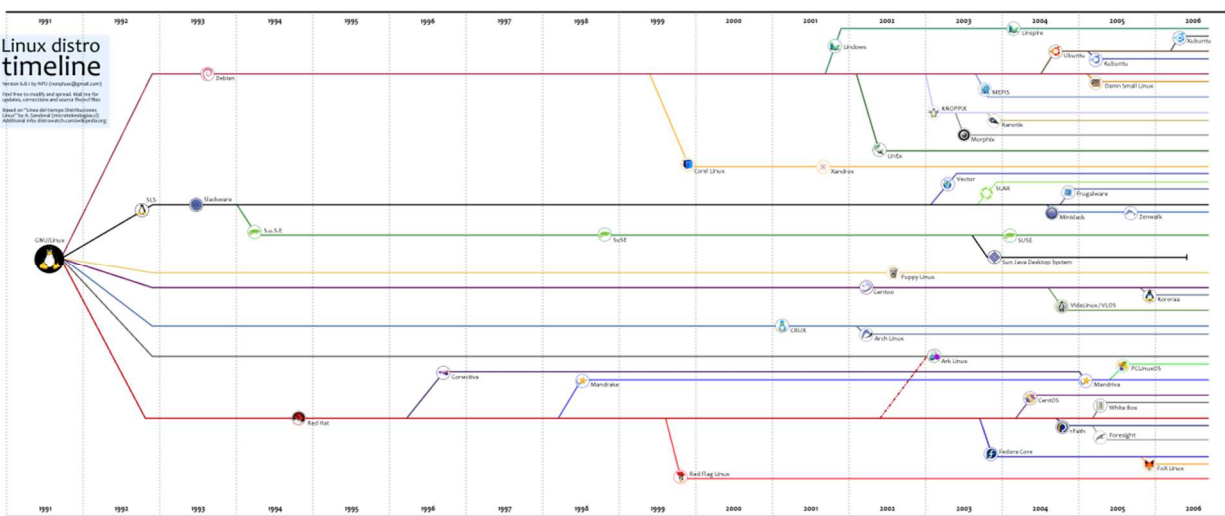
Related to these transaction and organizational costs is the means through which *coordination* is enacted within either market or firm. Within firms, organizational costs refer to the means by which production is completed: select agents with known talent sets toiling towards company goals handed down from management. The “cost of business” is the ability for that firm to establish rules for coordination that enable a group of known agents to complete tasks to the satisfaction of those who would engage that firm for a select purpose. The agents, in this case, are known. They were typically hired for the talents they had when they joined that organization rather than for the promise, creativity, or motivation they showed when they were still outside the door. For their continued successful participation in the organization they are typically rewarded monetarily, occasionally with praise or promotion. Any significant lapse in these agents’ abilities to complete tasks could typically mean that their role within that organization is forfeited; the agent must succeed to known standards or perish, and another agent will be brought on instead to strive towards those known goals. From this simple perspective, coordination largely entails the ability to create the hierarchy that establishes the goals (i.e., known projects), to maintain the staff to achieve those goals (i.e., known agents), and to create and adhere to the processes which ensure that those goals can be achieved (i.e., codified rules for coordination).

Distinct from these markets, firms, or other hybrid models, is the notion of commons-based peer production. In this form of production, individual agents address the problem of task selection not through external price signals or as the result of managerial decree. Rather, in this instance, production is the outcome of a decentralized process of information organization, relying on the distributed creativity of scattered communities completing both large and small tasks to contribute to a unified whole. In other words, task selection in peer-production is dependent on each individual agent’s capacity to both identify available tasks within that community and to accurately assess their own suitability to successfully complete that task. Through this decentralized model of production, collaboration is an emergent process extending from the community itself – there are typically no enforced roles. Hierarchy, if it exists at all, is largely flat. Coordination is represented only as the aggregation of tools, interfaces, opportunities, and individuals interacting within the system. And task selection, to say nothing of task completion, is at the behest of each individual agent within that community. Further, the types of production possible within these communities are manifold<sup>1</sup>. To clarify the capacity for peer-production communities to effectively organize and complete tasks, two exemplars are given and described below.

---

<sup>1</sup> For an idea of the breadth of current commons-based peer production projects, the P2PValue Project Directory (<http://directory.p2pvalue.eu/explore/cbpp-communities>) maintains a list of currently active projects along with their primary distinguishing characteristics.

First, we turn to the commons-based peer production of Free/Libre Open Source Software (FLOSS)<sup>2</sup>. Specifically, we start with the case of the Linux operating system<sup>3</sup> first announced via a Usenet post on August 25<sup>th</sup>, 1991. What started as a personal project, once released, was quickly adopted by a community of developers (Figure 1 provides a brief illustration of how the distribution was adopted and branched over time). In fact, since tracking began over 10 years ago, nearly 12,000 from over 1,200 different organizations have made contributions to the Linux kernel, with over 4,000 developers contributing between September 2013 and February 2015 alone, half of whom were first-time contributors<sup>4</sup>. While the issue of *what makes this project successful* will be covered more fully in Part I below, what is largely understood to be an inarguable claim is that the project is, in fact, successful. And while it still has a relatively modest 1.68% share of the desktop market, thanks to its prevalence in embedded devices and its 51.07% dominant share of mobile and tablet devices, it is clear that this model of production has been quite effective<sup>5</sup>.



**Figure 1. Timeline showing the multiple branches and individual operating systems (over 30 in all) that resulted from the release of the initial Linux kernel. Source: Wikimedia commons, Linux\_Distro\_Timeline.**

Our second example, and the site that will provide a case study for both the analysis of and design for commons-based peer production communities that will be the primary focus for the dissertation that follows, is Wikipedia, introduced in the section below<sup>6</sup>. As the primary theory and literature review will be included in Part I and beyond, the sections that follow are intended to provide a brief, high-level overview of the encyclopedia – what it is, how it started, and what that may mean for our evolving notion of community.

<sup>2</sup> For a more detailed explanation of the provenance and politics of the FLOSS designation, Richard Stallman provides a brief primer - <http://www.gnu.org/philosophy/floss-and-foss.en.html>. For further suggestions, also consult (Benkler, 2002, footnote 2).

<sup>3</sup> Another hallmark of the FLOSS movement: the source code for Linux is available from <https://github.com/torvalds/linux>.

<sup>4</sup> Statistics retrieved from the February, 2015 Linux Development Report, <http://www.linuxfoundation.org/news-media/announcements/2015/02/linux-foundation-releases-linux-development-report>

<sup>5</sup> Statistics retrieved from NetMarketShare, <http://marketshare.hitslink.com/operating-system-market-share.aspx?qprid=8&qptimeframe=M&qpcustomd=1>, retrieved August, 2015.

<sup>6</sup> Link to the Wikipedia main page, for the uninitiated: <https://www.wikipedia.org/>, retrieved August, 2015.

## PEER PRODUCTION OF THE SUM OF ALL HUMAN KNOWLEDGE: A BRIEF INTRODUCTION TO THE CURIOUS CASE OF WIKIPEDIA

In a Slashdot post from July 28<sup>th</sup>, 2004, Jimmy “Jimbo” Wales, cofounder of Wikipedia was quoted<sup>7</sup>:

*Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge. That’s what we’re doing.*

In what must now be a familiar refrain to researchers of online systems – Wikipedia is an online encyclopedia first launched in January of 2001. Running on the open-source MediaWiki platform, Wikipedia introduced users to a novel new form of collaboration and communication – an online encyclopedia that *anyone can edit*. Since its inception, Wikipedia has grown to be one of the most visited sites on the web, currently ranking number six in visits from the United States and number seven globally<sup>8</sup>, with over five million articles, over 25 million registered users, comprised of over 487 million edits by registered users and over 138 million edits by anonymous users in the English Wikipedia alone<sup>9</sup>. And while the rate of contributions has slowed since the encyclopedia’s explosive growth between 2004 – 2007 (Figure 2), it remains an invaluable resource to the millions who rely on the information presented on a daily basis<sup>10</sup>. Deferring a more detailed examination of the potential causes of, research into, and ramifications resulting from this decline to Part I, we return our focus to the larger questions of the *community* responsible for the creation and maintenance of such an information and cultural resource.

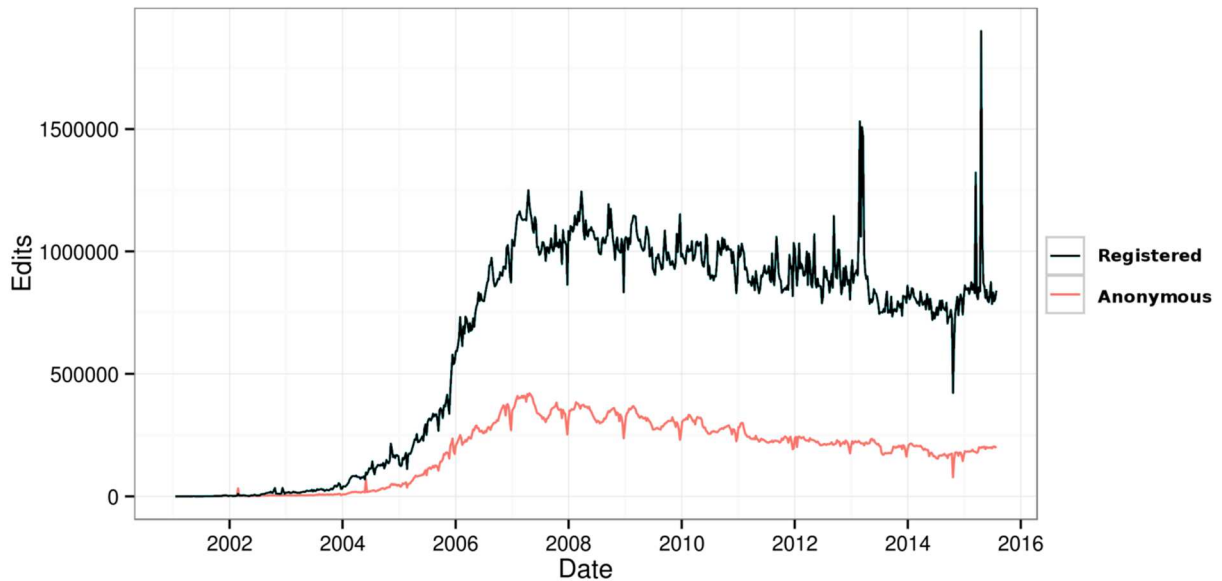
---

<sup>7</sup> Source: [https://en.wikiquote.org/wiki/Jimmy\\_Wales](https://en.wikiquote.org/wiki/Jimmy_Wales), retrieved August, 2015.

<sup>8</sup> Visit ranking provided from Alexa, <http://www.alexa.com/siteinfo/wikipedia.org>, retrieved August, 2015.

<sup>9</sup> All data was initially collected from the Wikipedia Toolserver (<https://toolserver.org/>) which was decommissioned in June 2014. Data is currently collected in aggregate from the Wikimedia Tool Labs ([https://wikitech.wikimedia.org/wiki/Help:Tool\\_Labs](https://wikitech.wikimedia.org/wiki/Help:Tool_Labs)) and made publicly available through an API more fully described in Part IV. The numbers above include all users and edits to the English Wikipedia through July, 2015.

<sup>10</sup> *Supra*, footnote 8.

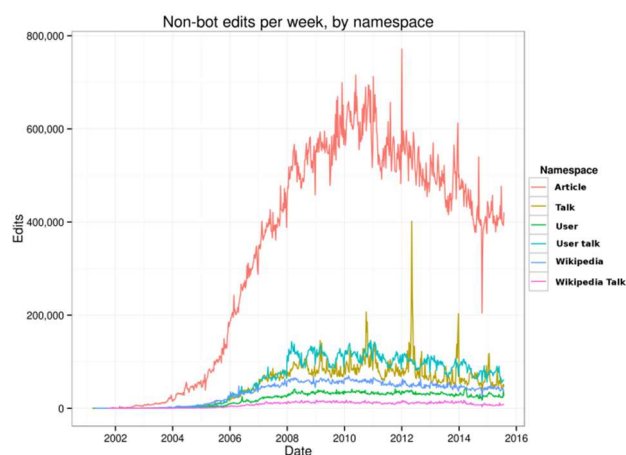


**Figure 2. Total edits per week to the English Wikipedia across all namespaces by both registered (blue) and unregistered (orange) users.**

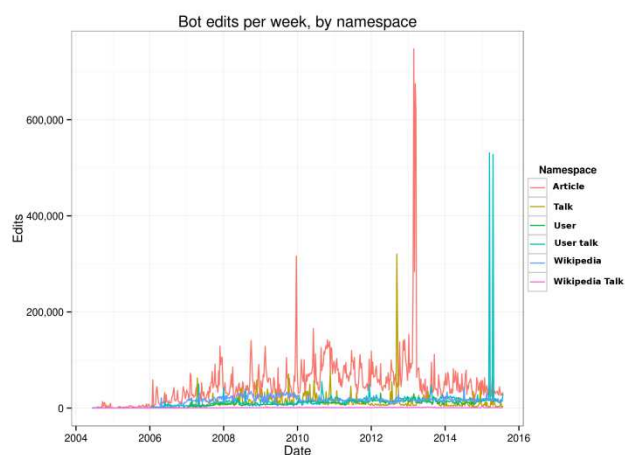
What type of community is present within those contributors to Wikipedia that facilitates the growth and maintenance of such an undertaking? Who are these individuals – largely acting without interest for monetary gain, largely contributing in their spare time – that are sacrificing the non-renewable resource of time for the sake of the encyclopedia, and by what means? Benkler (2002) states that this new model of production relies on “decentralized information gathering and exchange to reduce the uncertainty of participants. It has particular advantages as an information process for *identifying and allocating human creativity available to work on information and cultural resources*,” [emphasis added]. In other words, given the technology or collective cooperation to identify resources requiring human efforts, those resources, once identified, can be improved for the betterment of the community as well as the larger society in which that community resides. Along similar lines, Clay Shirky introduces the term “cognitive surplus” to identify both the need and the opportunity for those in society to take advantage of individual latent creativity (2010). Or, put simply, that if those in society spent the 200 billion hours watching television each year on any more productive pursuits, even those requiring only the most minute expressions of creativity, it is unimaginable that those hours spent in creative pursuits – the better utilization of collective cognitive surplus – would not have at the very least a modest impact in the enactment of information and cultural resources, returning to Benkler’s phrase<sup>11</sup>. Shirky goes on to clarify, “The cognitive surplus is not simply trillions of hours of free time

<sup>11</sup> To further support this claim, I once again refer to (Benkler, 2002): “Peer production provides a framework within which individuals who have the best information available about their own fit for a task can self-identify for that task.” The criteria for success, in this case, are first that visibility of available tasks is adequate to allow individuals to identify potential contributions and second, that there exists enough individuals within that community willing to make those contributions, both suppositions validated by previous contributions to the community (see Figures 2 through 4).

spread across two billion connected individuals. Rather, it is communal; we must combine our surplus free time if it is to be useful, and we can do that only when we're given the right opportunities," (2010, p. 97).



**Figure 3. User (non-bot) edits per week to the English Wikipedia, by namespace.**



**Figure 4. Bot edits per week to the English Wikipedia, by namespace.**

At a basic level, these “right opportunities” for contribution are two-fold. First, as the population of potential contributors to an information resource such as Wikipedia increases, the likelihood of any individual within that population identifying work that they are uniquely suited to completing also increases. And second, and following from the prior, as the *structure* of the community and the *tools available* to those potential contributors more effectively facilitate the identification of tasks available that are uniquely suited to each contributor, the likelihood of those potential tasks being identified and subsequently completed also increases. In the next section, we introduce some of those basic structures within Wikipedia.

## Structures of Collaboration: Projects, Problems, and Solutions

Group coordination in online spaces poses some unique challenges. Among them, it is typical that group members are not collocated, potentially with individual team members not even being located in the same time zone. Accordingly, communication among team members can be difficult and frequently only occurs in asynchronous fashion. Related, knowledge of other group members, their talents, their desires, and their schedules and offline interests are frequently unknown. Policies and social norms are frequently only gleaned through immersion in that group – a process that can only be accomplished through time, increasing the barrier to entry for new participants who may otherwise have been willing contributors. Rules and governance structures frequently exist, but in a similar vein to the prior issue, they are often impenetrable to new members, ensuring an entrenched minority controls the rules for the majority. Given these myriad problems and difficulties, it is indeed remarkable that these types of projects exist, let alone succeed. However, inarguably, Wikipedia has. One of the reasons for this success is the WikiProject.

Wikipedia defines a WikiProject as “a group of contributors who want to work together as a team to improve Wikipedia.”<sup>12</sup> The scope of these projects can vary widely - from cats to technology to military history - WikiProjects exist for a variety of subject matter and task centered areas. Beyond these topic-focused areas task-oriented projects also exist, such as those dedicated to identifying and resolving potential copyright violation, projects to monitor and remove articles not deemed “encyclopedic,” or projects focused on reaching out to potential new editors in an effort to attract a greater number of contributions from a wider range of the population. And while overall levels of contributions appear to have been on the decline since 2007, as previously observed, the number of WikiProjects has continued to grow, with more “Wikipedians” identifying with these projects and engaging in focused, directed editing work under the scope of each one. In fact, as of August, 2015, there have been over 2,300 WikiProjects created across a vast spectrum of both topic- and task- related areas<sup>13</sup>.

We argue that these WikiProjects, these open structures of distributed collaboration and the ongoing activity within them, are among the primary means of analyzing, theorizing, and designing for the larger community; that they are a means of providing the “right opportunities,” returning to Shirky, for these distributed team members to “scour larger groups of resources in search of materials, projects, collaborations, and combinations...” (Benkler, 2002, p. 376) to which those community members may contribute. The sections that follow present the means through which these statements may be evaluated, the theoretical foundations by which they may be expanded upon, and the justifications for design interventions that may facilitate the ongoing interactions within these myriad teams existing within the complex and emergent environment of Wikipedia.

## **STRUCTURE OF THIS DISSERTATION & RESEARCH GOALS**

Continuing from the above, this dissertation aims to answer three primary research questions, enumerated first and expanded upon below:

- 1) Through what means can we best analyze coordination in commons-based peer production communities to more effectively inform future research and support theory-driven design? Given that successful intervention requires accurate interpretation of activity in a massive, complex system, the means by which that interpretation occurs must be appropriately grounded and adequately conveyed to better facilitate community interaction with future design.
- 2) Given the complexity of coordination and information transfer within distributed peer production communities, how can the prior interactions identified and ensuing analysis contribute to our theoretical understanding of the WikiProjects under observation?
- 3) How can this formal, theoretical interpretation of distributed interactions inform design to best facilitate ongoing interactions and foster future growth?

---

<sup>12</sup> Source: <https://en.wikipedia.org/wiki/Wikipedia:WikiProject>, retrieved August, 2014.

<sup>13</sup> Data collected from a custom API from information available from the MediaWiki Tools Lab servers. Details on this API, including the means to access or replicate it, are included in Part IV.

Accordingly, this dissertation will be presented in five parts, with each building upon the last, each providing a layer of mortar to the theory that is being built, examined, and evaluated. The intent of this structure is to allow the reader of this dissertation the ability to recreate, as fully as possible, the steps leading to the original impetus for this work, the initial analyses to actively explore the communities being studied, the construction of a theory to more fully describe those communities, and finally, the means by which that theory was applied. This will entail an explication of the tools required to both ask the proper questions about behavior within modern commons-based peer production communities, the means by which those behaviors are analyzed and explained, and finally, the creation of a theory to more effectively design for those communities.

Part I of this dissertation will highlight prior research that has been completed with regards to commons-based peer production communities, including both their provenance, their importance, and the means by which they can be differentiated from traditional market-based, firm-based, or hybrid organizations. This includes a more detailed introduction to the work practices of individuals active within the community, the emergent groups and collaborations that arise through the enactment of this activity, and an overview of the research that has been done on both. Part I concludes with a detailed introduction to prior work that provides the theoretical foundation for the research that this dissertation aims to address.

Stemming from the discussion of the structure, activity, and analysis within commons-based peer production communities, Part II introduces two studies that aim to cement these processes in the current work. Each of these studies, in turn, provides concrete examples of both the hurdles inherent in studying the vastly shifting realities of a voluntary contributor driven community as well as the means by which these hurdles can be addressed. Each of these studies extends from a simple presupposition: in an environment that is wholly represented and experienced as a collection of technologies, from the frameworks that underlie the façade to the designs and algorithms that comprise and ultimately enrich its expression, the ability to *measure activity* in these communities affords the ability to *design activities* within them that will, ultimately, provide a theoretical foundation to more effectively *understand activity* within that system<sup>14</sup>. Each study presented in Part II, then, aims to more fully expand upon these larger themes.

Once these foundations are established, terms are defined, and a common language is introduced, Part III of this dissertation expands upon these prior examinations towards the primary contribution of this dissertation: a theory of coordination for commons-based peer production communities. This theory, borne out by the perspectives gleaned from these prior explorations, extends from a grounded examination of interview transcripts with a broad spectrum of individuals active within this community. Through their experiences, both those desirable and those frustrating, the theory of coordination presented provides a meaningful lens through which to better identify and understand instances of *information transfer* within the systems they are interacting with, providing a unique tool for designers

---

<sup>14</sup> Variations on this idea are intriguingly expanded upon in (Geiger, 2014) and more directly explored in (Geiger & Ribes, 2011). From the latter, Geiger & Ribes state explicitly: “documentary traces are the primary mechanism in which users themselves know their distributed communities and act within them.” In other words, the technological platform through which interactions take place is not simply a mediator for those interactions; it is the sole means by which identity, communication, and information transfer take place.

of interactive systems as well as the users of those systems to better understand the means by which myriad interactions occur, the multiple ways in which those interactions may fail, and ultimately may suggest potential solutions through which those prior risks identified may be mitigated.

Following from the introduction of a modified theory of coordination for commons-based peer production communities, Part IV of this dissertation introduces a novel prototype application which, utilizing that prior theory to identify potential breakdowns of information transfer, addresses and improves the means by which distributed voluntary virtual team members can coordinate their actions. This tool, the Virtual Team Explorer, illustrates the trajectory of a practical example in which empirically derived data and analyses (Part II) yields improved understanding of community behavior, which in turn provides the building blocks of a theory of coordination to describe that behavior (Part III). Part IV concludes with a discussion of the potential ramifications for such theory-driven design interventions, outlining the means through which such a theory of coordination can be used as a lightweight methodology to better understand large scale user behavior and to, ultimately, design more effective tools for those users.

Finally, Part V of this dissertation provides a conclusion to the work presented, including a discussion of the merits of such theory-driven design approaches on both analyses of and design for such systems, as well as a discussion of the importance for such approaches to continually improve and refine the types of increasingly complex interactions that are possible within them. Further, moving beyond Wikipedia such approaches provide a lightweight means to quickly identify moments of *relevant* information transfer, providing system operators and designers a theoretically grounded means of identifying potential breakdowns within that system. Ultimately, I argue that the work presented here illustrates the provenance, the purpose, and the development of a theory of coordination for commons-based peer production communities. Further, the creation of the Virtual Team Explorer is presented as a proof-of-concept of those theory-driven ideals, indicating the many ways in which a deeper understanding of these complex organizations may be utilized, as well as indicating how such an approach may be adopted as a lightweight methodology towards better understanding users' interactions and, very likely, their potential frustrations.

In short, this dissertation aims to communicate four primary goals. First, it will provide a concrete foundation regarding the composition of the commons-based peer production community under study, including an introduction to both the historical and theoretical underpinnings of how that community has been examined previously. Second, it will introduce the practical means by which that community may be better understood and examined, including an exploration of both the means of examination and the myriad tools and methodologies available to the examiner. Third, it will introduce a theory of coordination for commons-based peer production communities based on the experiences of its actual members, driven by the necessity to ground design for such communities in actionable observations. And fourth, it will introduce a prototype application intended to provide a meaningful intervention into that community, to address and potentially ameliorate problems inherent in large scale distributed collaborations. In other words, this dissertation aims to describe the process of answering the following questions with regard to one specific commons-based peer production community: How do we measure it? How do we build for it? How do we justify design interventions? And what might such an intervention look like? Ultimately, the primary contribution of

this dissertation is an attempt to answer these questions, basing such explanations in both the examination of and interaction between the members of that community, over the course of years for the latter, and over the course of its entire existence for the former.

### **A Brief Note on Authorial Voice**

As the majority of the work presented in the following sections was completed in collaboration with others, and as all research is, in essence, comprised of an accumulation of knowledge from the distributed efforts of those that come before you, the first person plural will primarily be used to describe the work completed. No research exists in a vacuum, and this work is no exception. When discussing the dissertation itself, the first person singular pronoun will be used to allow for a more clear distinction between the analyses that comprise the majority of the work and the discussion that stems from the work itself.

# Part I – Overview of Prior Research: Foundations for Development

*Knowledge [is not] a matter of getting reality right, but rather... a matter of acquiring habits of action for coping with reality.*

(Rorty, 1992)

*Under their standard interpretations, the cards serve exclusively as paint samples. They are mere instances of the other properties, but telling instances of the colors. A symbol that is a telling instance of a property exemplifies that property. It points up, highlights, displays, or conveys the property. Since it both refers to and instantiates the property, it affords epistemic access to the property.*

(Elgin, 2008)

---

The goal of this chapter is to provide background context for the studies and discussion that will follow. This includes the varied ways in which *participation* is measured in this distributed community of contributors, the means through which group *identification* and *membership* have been encouraged and established, as well as research on how distributed teams within WikiProjects *function*. Extending from this prior research, I also explore different ways through which behavior in Wikipedia has been theorized, focusing in particular on analyzing WikiProjects from the perspective of Benkler's commons-based peer production communities. Finally, I introduce Coordination Theory, the lens through which this dissertation will initially classify interactions within WikiProjects, and which will ultimately be expanded to provide a more incisive tool to effectively inform theory-driven interventions into this and similar communities.

## 1.1 UNDERSTANDING WIKIPEDIA

Continuing from the earlier introduction to Wikipedia and WikiProjects, the following sections will provide more background into how prior research has been conducted within these communities and provide context for the research that is to follow in this dissertation. Much of the work that follows involves, first, understanding and interpreting needs and interactions within distributed teams and, second, designing theory followed by technology to better describe and suit those needs. Accordingly, this section begins with an exploration of how participation and value has been measured in Wikipedia, with particular focus on how distributed activities within commons-based peer productions may be interpreted. Following, I introduce the idea of *coordination* as a meaningful sensitizing concept in this research, specifically with regard to commons-based peer production communities. Finally, by adopting coordination as a unifying concept in approaching analytical problems within peer production

communities, I argue that the interpretation of that research will be more concretely grounded in the nature of those communities, providing a common language to effectively approach theory and design.

### 1.1.1 Understanding Wikipedia: How Measurement Shapes Participation and Value

A primary attribute of commons-based peer production communities introduced above is the decentralized nature in which they typically operate. Without formal hierarchy or externally imposed goals, the nature of assessing the value of individual agents within a community is challenging. Accordingly, the following examples highlight a few of the varied means that researchers have attempted to unpack interactions within Wikipedia. To wit, the following interaction between Jimmy Wales, cofounder of Wikipedia, and another contributor from May, 2005 regarding the question of who truly contributes the most value (i.e., content) to Wikipedia<sup>15</sup>:

Sj wrote:

*The swarm does the bulk of the writing, especially finding and providing current facts, starting new articles, and adding neglected POVs. The roving groups are sensitive to dozens of policy pages, and implement them as they rove... they also take on large projects, one at a time, and try to implement certain changes across thousands of pages at once.*

Jimmy Wales responded:

*[...]*

*My own research indicates the opposite [...]*

*My research (conducted in December) showed that half the edits by logged in users belong to just 2.5% of logged in users. It would be extremely interesting to run tests to compare “edit dispersion” for new articles, old articles, heavily edited articles, highly watched articles, heavily trafficked articles, etc.*

*A deeper understanding of all these issues can have some interesting implications for us in terms of understanding certain policy issues.*

Throughout this thread, Wales opposes the notion that Wikipedia is an example of the “wisdom of the crowds,”<sup>16</sup> in which the disparate masses, the swarm, acts in an uncoordinated fashion to create something wholly greater than the sum of its parts. Instead, he states:

*[...] I'm not convinced that ‘swarm intelligence’ is very helpful in understanding how Wikipedia works -- in fact, it might be an impediment, because it leads us away from thinking about how the community interacts in a process of reasoned discourse.*

---

<sup>15</sup> The complete thread in which this conversation took place is at <http://marc.info/?t=111502471000003&r=1&w=2>, retrieved August, 2015.

<sup>16</sup> Referring to the then-just released *The Wisdom of the Crowds*, by James Surowiecki, 2005.

The facts he presented were accurate; however, the presupposition in the interpretation of those facts was not based on a perspective that aligned with actual behavior at that time. To test this theory that the majority of the encyclopedia was written by “a community... a dedicated group of a few hundred volunteers,” Aaron Swartz looked at a sample of articles to determine *who* in fact could be deemed most responsible for creating the content within Wikipedia: the previously mentioned dedicated community of few, or the swarm of the whole. However, instead of using revision counts per user (i.e., the number of times each user committed a single change), he decided to measure contributions by the number of characters added in each revision (i.e., users who added a paragraph would be more impactful than those who fixed a punctuation error). In September, 2006, Swartz wrote<sup>17</sup>:

*But when you count letters, the picture dramatically changes: few of the contributors (2 out of the top 10) are even registered and most (6 out of the top 10) have made less than 25 edits to the entire site. In fact, #9 has made exactly one edit – this one! With the more reasonable metric – indeed, the one Wales himself said he planned to use in the next revision of his study – the result complete reverses.*

*[...] For example, the largest portion of the Anaconda article was written by a user who only made 2 edits to it (and only 100 on the entire site). By contrast, the largest number of edits were made by a user who appears to have contributed no text to the final article (the edits were all deleting things and moving things around).*

The discrepancies between these findings, and their inclusion in this dissertation, are not the result of a mismanagement of statistics. Rather, they are the result of a disagreement of *construct validity* (Shadish, Cook & Campbell, 2002). In short, if the question is, “who are primarily responsible for the sum total of the content on Wikipedia,” data collection must be completed in a manner where the data collected can effectively answer that question (regardless of prior hypotheses). For the sake of brevity, we will sidestep the semantic argument of defining individual terms within the preceding question – while it is true that “primarily responsible” has not been effectively codified, and variations of interpretations of “content” exist that could make either finding accurate, as the understood goal was to verify, from the reader’s perspective, what portion of what was read was contributed by a member of the “swarm” versus a core contributing member.

Finally, a 2007 study led by Kittur (Kittur et al., 2007) tackled this question as well. In their study, Kittur stratified contributors by both explicit (administrator versus non-administrator) and implicit (edit counts) activity and analyzed each group’s contributions over the (then) life of Wikipedia. Given this more large-scale analysis of activity (Swartz’s analysis was completed using only a small handful of articles), their research ultimately found that while the “elite” users (the most active editors and the administrators) had a greater impact on the amount of content added in the period between 2001 – 2004, in the years following it was the “common” users who were most responsible for the majority of the work completed, regardless of whether the work was measured by edits or content changed. Further, they replicated this finding on the social bookmarking site del.icio.us, suggesting that these types of shifts may occur in all contributor communities, likening it to a “rise of the bourgeoisie” (Kittur et al., 2007).

---

<sup>17</sup> The complete blog post in which this quote was found is at <http://www.aaronsw.com/weblog/whowriteswikipedia>, retrieved August, 2015.

What these examples ultimately aim to highlight is the complexity through which conclusions can be drawn in commons-based peer production communities. In an encyclopedia in which anyone *can* edit, it remains an inherent possibility that at any given time, everyone *may* be editing. When the deadline for adding improved content is both *right now* (in that every moment an article contains sub-par content is anathema to the goal of the encyclopedia) as well as *whenever you get to it* (in that those contributing that content are doing so of their own volition, on their own time, at the behest of their own motivations)<sup>18</sup>, any conclusive quantitative measure of progress or relative value is difficult to establish.

The implications of this issue are clear. Regarding policy, if the most casual contributors are predominantly responsible for the content contributed, enacting norms that specifically support those users would likely improve their longevity with the system. Related, regarding onboarding, ensuring that early experiences with the platform are positive will increase the likelihood that those responsible, in aggregate, for the most content will continue to contribute. Regarding design, while it may be important to provide tools for elite users early on to structure the content that underlies the community, also ensuring that there are lightweight means of supporting valued contributions from the masses will likely also facilitate continued growth. The implications for *measurement* are equally clear: if analysis is completed without sensitivity to the nature of commons-based peer production communities, the prior implications will be mistaken. The risk will be that policy is created without considering the community, design will favor the elite, and new users may be automated out of existence. In short, to effectively support a community it is necessary to accurately understand that community. And with regard to understanding commons-based peer production communities in particular, it is necessary to frame analysis and ensuing interpretation utilizing a lens that facilitates understanding the community *in situ*, allowing researchers to derive accurate and actionable insights from that analysis. We argue that such a lens is *coordination*.

### **1.1.2 Prior Conceptions of Coordination in Wikipedia**

Given that Wikipedia has by most metrics been largely successful since its launch in 2001, it should be unsurprising that there has been a large amount of research into exactly how that success was achieved. This section aims to highlight some of that research, focusing on how prior research has, first, framed the notion of coordination, and second, through what methods did that research analyze how coordination was enacted.

#### **Operationalizations of Coordination**

Research led by Kittur (Kittur et al., 2007) aimed to identify how non-direct work has increased throughout the growth of the encyclopedia. In this study, “coordination costs” were defined as “excess work in the system that does not directly lead to new article content.” In other words, coordination was defined as a subset of edits that were *not* directly to article content, including those to article talk pages, user and user talk pages, and pages in the Wikipedia

---

<sup>18</sup> Example from a personal conversation, OpenSym 2013.

namespace, where most Wikipedia policy pages reside, as well as WikiProject pages<sup>19</sup>. Findings from this study indicated that the coordination costs of creating content in Wikipedia had increased throughout its existence, with a greater amount of efforts being applied to the *indirect work* of maintaining the encyclopedia. A separate study led by Kittur (Kittur, Pendleton, & Kraut, 2009) sought to identify whether group identification (i.e., becoming a “member” of a WikiProject by including their editor name on a project member list) would have a positive impact on overall activity, activity within the scope of the WikiProject, coordination work, and good citizenship behaviors (e.g., fighting vandalism), ultimately finding that group identification did indeed increase these behaviors. In this study, “coordination work” was even more narrowly defined as “whether edits were made to article pages directly (work type = 0) or to the discussion pages (work type = 1),” indicating that the proxy for coordination, in this case, was the number of edits each contributor made to an Article talk page. Another study led by Kittur (Kittur & Kraut, 2009) sought to examine how coordination among editors would impact article quality. In this study, coordination was framed in two ways: *implicit coordination* was defined as a single editor taking on the majority of the work for any given article, while *explicit coordination* was defined as communication around the development of that article on its Talk page. Ultimately, they found that both types of coordination led to an increase in article quality; however, implicit coordination was particularly effective when higher numbers of contributors were present while explicit coordination was more effective with lower numbers of contributors.

One study led by Viégas (Viégas et al., 2007) sought to, in part, examine exactly what types of coordination were occurring on a sample of 25 article talk pages. In this study, the authors manually coded all posts to article talk pages within their sample along 11 dimensions. These dimensions explicitly included “Requests/suggestions for editing coordination,” defined as “Postings that help users plan editing activity,” along with other categories such as “Requests for information,” “References to Wikipedia guidelines and policies,” “Requests for peer review.” Ultimately, the authors found that “Requests/suggestions for editing coordination” was the predominant post type, with an average of 58.8% falling within that category, distantly followed by “Requests for information” with an average of 10.2%<sup>20</sup>. A separate study led by Morgan (Morgan et al., 2013) similarly aimed to classify 788 thread-starting posts to active WikiProject talk pages using a refined and expanded version of the codebook used by (Viégas et al., 2007). The Morgan et al. study contained two categories explicitly referencing coordination: “REQ-COORD-ART,” defined as a request or proposal for coordinated editing of articles, and “REQ-COORD-NONART,” defined as a request or proposal for coordinated editing of all non-articles (e.g., asking for assistance adding articles to categories). This study found that, contrasted with earlier findings by Viégas et al. (2007), both coordination

---

<sup>19</sup> It is unclear if edits to pages in the Wikipedia talk namespace were included (where most debate and discussion surrounding policy would occur), or edits to pages in other namespaces where coordination work may be completed (e.g., templates, for instance).

<sup>20</sup> While some prior studies classified *all* posts to article talk pages as coordination work, Viégas et al. utilized a codesheet to classify talk page posts with only one category explicitly labeled as “coordination.” One could argue that interpretations of coordination may also include other categories in their codebook such as “Requests for information” or “Requests for peer review” as well, to a greater or lesser extent, however that level of analysis is outside the scope of this review.

categories were only coded for 4% and 3%, respectively, with REQ-OPINION being the primary category identified (defined as “Requests advice, opinion or informal feedback, or attempts to gauge local consensus around an idea”)<sup>21</sup>.

A separate study led by Morgan (Morgan et al., 2014) took as its focus the nature that WikiProjects have changed over the life of Wikipedia. In this study, the authors distinguish between *conventional* WikiProjects, defined as “WikiProjects that are generally scoped around an encyclopedic topic and focus on coordinating article editing tasks,” and *alternative* WikiProjects, defined as any project that deviates from the “norm in stated goals, scope or primary tasks.” The distinction here departs from the assumptions in the above studies. Rather than frame coordination as all work occurring on a WikiProject or article talk page, the authors identified task-focused alternative projects as a proxy for coordination within Wikipedia. For instance, while WikiProject Cats would be a conventional, topic-focused project, projects without a focus on a specific topic would be categorized as alternative or task-focused, such as WikiProject Articles for creation. Ultimately, they found that while participation in conventional WikiProjects roughly matched the decline in overall activity in Wikipedia as a whole, the amount of participation in alternative WikiProjects (i.e., the number of edits to those project pages, sub-pages, and corresponding talk pages) had actually increased 57% in the period from 2007 to 2012.

Beyond these studies with explicit operationalizations of *coordination*, several studies have been completed that would facilitate interrogating issues of coordination within projects. For example, on study led by Cosley (Cosley, Frankowski, Terveen, & Riedl, 2007) in part examined the impact of an intelligent task routing system’s ability to direct the editing behavior of users who received its task recommendations. This system, called SuggestBot, was shown to have a positive impact on directing efforts, ultimately showing that their task suggestions were edited about four times more than random task suggestions. Although this site makes no explicit reference to coordination, we include it in this review as an example of a study that has the capacity to directly impact contributors’ perceptions of task availability, a critical requirement for contributing to commons-based peer production communities, as will be discussed further in the sections to follow.

## Theorizing Coordination in Wikipedia

Using a different approach, one study led by Bryant (Bryant, Forte, Bruckman, 2005) utilized the lens of Legitimate Peripheral Participation (Lave & Wenger, 1991) to explore how contributors progressed from casual novices to expert Wikipedians, explaining their results from the perspective of the Activity Theory framework (Engeström, 1999). Through qualitative analysis of a series of semi-structured interviews, Bryant et al. found that as each of their interviewees progressed from newcomer to expert (i.e., from peripheral to Wikipedian), their nature of their *activity*<sup>22</sup> was also dramatically changed. For example, editing for novice users was typically ad-hoc; they would identify an article on a subject they were familiar with and, if they were comfortable, would complete and commit their change. During this process, the authors report that none of the novice contributors used any of the tools at

---

<sup>21</sup> Possible causes for the discrepancy between findings by Viégas et al and Morgan et al include the former was categorizing all posts to article talk pages while the latter categorized only thread-starting posts on WikiProject talk pages, the refinement of categories that occurred between the studies, or differences between individual coders. However, as the focus of this portion of the review is on framings of coordination in prior research, exploration of these possibilities will be left to the reader.

<sup>22</sup> “Activity” here referring to the more rigorous form of the term as defined in Activity Theory.

their disposal in the course of their first edits, such as referring to the article talk page or viewing the revision history of the article. Conversely, the authors found that experienced Wikipedians, while they would still edit individual articles, were in general far more concerned about the health of Wikipedia as a whole. Further, through the course of their activities with Wikipedia, these more experienced users were also reported to rely far more on the tools available to them, such as the watchlist, recent changes pages, as well as talk pages, which the authors report are the “most frequently used communication channel on Wikipedia” (which to some extent substantiates prior operationalizations of coordination within the community). In short, the authors found that by taking a more longitudinal look at the development of individual Wikipedians over time from the perspective of Activity Theory, behaviors progressed from isolated production (e.g., simple article editing with little communication) to ownership and mediation (e.g., task-focused monitoring of articles, dispute resolution, or community *coordination*).

Another study led by Viégas (Viégas, Wattenberg, & McKeon, 2007) analyzed the Featured Article process<sup>23</sup> through the lens of commons governance (Benkler, 2002) and four of Ostrom’s eight principles for self-organizing communities (Ostrom, 2000). These principles include<sup>24</sup>:

- 1) Congruence between rules and local conditions: rules which are based upon the characteristics of the current environment as opposed to global or generic, or the congruence between the benefits received from individual effort and the input required to keep them (2),
- 2) Collective choice arrangements: the ability for those impacted by local rules to modify those rules (3),
- 3) Monitoring: individuals responsible for oversight who are accountable to their local community (4), and
- 4) Conflict resolution mechanisms: the ability for community members to easily address conflicts in an accessible forum (6).

Ultimately, Viégas et al. conclude that one potential explanation for the success of the Featured Article process in Wikipedia is a result of its emergent alignment with these design principles facilitated by the structural characteristics of commons-based communities. Specifically, the capacity for Featured Article reviewers to impact the criteria by which Featured Articles are selected, through both process and technology, indicates alignment with principle 1. The capacity for reviewers (as well as all Wikipedians) to participate in consensus building and policy creation indicates alignment with principle 2. The capacity for Wikipedians to monitor changes through watchlists and templates indicates alignment with principle 3. And the highly public and visible nature of Wikipedia’s policy and consensus environment indicates alignment with principle 4. In short, Viégas et al. utilized Ostrom’s principles as a post hoc validation to explain the success of the Featured Article process in Wikipedia.

Extending this prior research by Viégas et al., two studies led by Forte (Forte & Bruckman, 2008; Forte, Larco, Bruckman, 2009) examined the emergent governance mechanisms within Wikipedia, also through the lens of

---

<sup>23</sup> This refers to the codified process through which articles are nominated, reviewed, and potentially promoted to Featured Article status, designating the highest quality content within Wikipedia. More information is at [https://en.wikipedia.org/wiki/Wikipedia:Featured\\_article\\_candidates](https://en.wikipedia.org/wiki/Wikipedia:Featured_article_candidates), retrieved August, 2015.

<sup>24</sup> As Viégas et al were only using four of the original eight design principles, the numbered list refers to the order each principle was used in their research, while the parenthetical number after each principle refers to the original numbering as identified by Ostrom (2000).

Ostrom's principles (Ostrom, 2000). In contrast to Viégas et al., Forte utilized all eight principles, with the remaining being<sup>25</sup>:

- 5) Clearly defined community relationships: enabling participants to “know who is in and who is out of a defined set of relationships and thus with whom to cooperate,” (Ostrom, 2000), or clear boundary rules (1),
- 6) Graduated sanctions: to ensure that any sanction is contextually appropriate and commensurate with the offense (5),
- 7) Local enforcement of local rules: the group must be flexible to the need for evolutionary adaptation as context, technology, or the needs of its members change (7), and
- 8) Multiple layers of nested enterprises: when the scope of resources is larger, governance activities should be in multiple layers of nested enterprises (8).

Similar to Viégas et al. (2007), these studies found that Ostrom's principles were largely accurate descriptors of the emergent governance environment and community structure within Wikipedia, also concluding that this parity between design principles and community could be responsible in part for the success of that community. One notable exception that Forte acknowledged was the inability of WikiProjects to enforce their own policy (i.e., while the scope of WikiProjects may increase the modularity of the community and allow a more structured distribution of labor, without explicit ownership or control over activity within the scope of each project it may eventually become difficult to avoid conflicts of quality or norms between projects). Finally, Forte et al. (2008; 2009) attribute much of the success of the larger community to the presence of “old-timers” responsible for maintaining and communicating community norms within the various projects, committees, and conversations that they contribute to. The Ostrom corollary to this last point would be principle one (principle five by Forte's numbering); in order for these old-timers to have a positive impact by this interpretation it would be required that other community members acknowledge or are aware of that seniority and to act accordingly.

## **Researching Wikipedia: Coordination in Context**

Consider how a polished and complete article is arrived at in Wikipedia. First, the article is created. It is possible that the first revision was already perfect, with structure, pictures, citations, and links to other relevant articles already in place. It is more likely that the article went through several revisions before that polished whole was achieved, with many individual contributors each piecing together new bits of content, improving language, or arguing the relative merits of including one piece of information over another. For instance, consider the article on the 2013 Rosario gas explosion<sup>26</sup> that was comprised of only two complete sentences when it was created on August 7, 2013, pictured below:

---

<sup>25</sup> Once again, the numbered list refers to the order each principle was used in Forte et al, while the parenthetical number after each principle refers to the original principle as identified by Ostrom (2000).

<sup>26</sup> This specific article was selected as it was “today's featured article” on the front page of the English Wikipedia on August 6, 2015.



**This article documents a current event.** Information may change rapidly as the event progresses, and initial news reports may be **unreliable**. The last updates to this article **may not reflect** the most current information.

Ten people have been killed and several injured in a gas explosion in the city of [Rosario, Santa Fe](#) in Argentina. Some candidates in the [Argentine legislative election, 2013](#) have suspended their campaigns.

## External links

- [report](#)

*This article is a stub. You can help Wikipedia by expanding it.*

**Figure 5. The first revision of the Wikipedia article “2013 Rosario gas explosion”.**

Between its creation and August 6, 2015, the article was edited another 254 times by 51 different users, including seven anonymous users, six bots, and 38 registered users. Of the users who contributed to the article, only two edited the article more than 10 times. And most notably, less than one year after the article was created it was promoted to Featured status, a distinction denoting the article as “one of the best articles produced by the Wikipedia community.”<sup>27</sup>

What is noteworthy is how this one example aligns with observations from prior research. As reported by Kittur & Kraut (2008), a larger number of editors only increases the quality of an article if there is a centralization of the work being completed on the article (i.e., implicit coordination through work concentration). In the case of the 2013 Rosario gas explosion article, this centralization took the form of a single user making the majority of edits as the article quickly expanded – 158 edits total over the lifetime of the article. Between August and November, 2013, the user “Cambalachero” expanded the article from 1,552 bytes to 29,469 bytes, adding structure, interwiki links, and increasing the number of references from 0 to 44. This user then shepherded the article through the GA process as well as the FA process<sup>28</sup>. Each of the contributions from this editor were made largely independently; while there were 12 non-bot users who posted to the article talk page, eight out of 25 edits were made by Cambalachero, and those were largely to add category and project information to the parent article. So, in this instance, we can interpret *implicit coordination* as a single contributor completing the majority of the work in silence and isolation.

This interpretation is problematic, however, examining the creation and development of the article through the previous theoretical perspectives. Lacking any significant interaction among editors there is very little opportunity to argue for the development of this or other editors that interacted on the page, no evidence for an increase in the *zone*

<sup>27</sup> Source: [https://en.wikipedia.org/wiki/Talk:2013\\_Rosario\\_gas\\_explosion](https://en.wikipedia.org/wiki/Talk:2013_Rosario_gas_explosion), retrieved August, 2015.

<sup>28</sup> The GA process refers to “Good Articles” in Wikipedia, articles that have been nominated and meet the Good Article criteria. More details are available from [https://en.wikipedia.org/wiki/Wikipedia:Good\\_articles](https://en.wikipedia.org/wiki/Wikipedia:Good_articles), retrieved August, 2015.

of proximal development<sup>29</sup>, and little support for an emergent social structure to facilitate the continued development of the article. While Ostrom's principles provide modest descriptive power for the *structure* of the article and related talk page, the level of interaction present does little to support any actionable intervention to improve coordination within those pages.

Given these problems and the research reviewed above, we argue that a more clear conception of *coordination* in commons-based peer production communities is required. Specifically, we argue that a combined means of analyzing coordination should provide actionable information (i.e., prescriptive data) that could lead to design interventions as opposed to the above theoretical studies which primarily *tested* theory against existing activity, i.e., post hoc validation of theory in an already successful community. Ultimately, we are making the claim that more finely honed theoretical perspectives are required within commons-based peer production communities to effectively direct design research and interventions. In the sections that follow we introduce the foundations of such a theory, while the chapters that follow in this dissertation will provide the provenance, creation, and testing of that theory.

## 1.2 COLLABORATION IN COMMONS-BASED PEER PRODUCTION COMMUNITIES

While the Introduction provided a brief overview of commons-based peer production communities including distinguishing features between these new methods of production and the markets and firms which preceded them, the following section aims to provide a more detailed overview of the structure and theory behind these communities.

Benkler (2002) describes the sustainability and organization of large-scale peer production communities through two analytic moves. The first move, he argues, is the proposition that a wide array of motivations drive individual behavior. Within this broad spectrum of motivations it is reasonable to argue, given a large enough population, there will exist individuals who value social-psychological rewards more highly than monetary rewards. Second, given any activity that can be broken into reasonably achievable sub-activities, the motivation to get any single individual to complete that sub-activity will be commensurately reasonable. From this understanding, Benkler states that successful peer production communities will share three primary characteristics. First, activity must be modular. To maintain parity between the motivation of individual members within a community and the motivation required to complete the activities necessary to maintain the health and function within that community, those activities will need to support disaggregation, ensuring that the diversity of motivation, availability, and talent present within the community matches the scope, complexity, and difficulty of tasks requiring completion. Second, the granularity of those sub-activities, or modules, needs to be sufficiently fine-grained to support contributions from the largest spectrum of community members – the more modules that can be completed by the greater portion of the

---

<sup>29</sup> As referenced in (Bryant, Forte, Bruckman, 2005), indicating the capacity to complete any activity given both interior knowledge and capabilities *combined with* exterior tools or networks.

population, the greater the likelihood that each module may be successfully completed<sup>30</sup>. And third, once individual modules have been completed, there must exist a mechanism for low-cost quality control and integration of that module back into the whole. In short, these characteristics describe the capacity of a project to support disaggregation (breaking activities into reasonably sized sub-tasks) and aggregation (reintegrating those subtasks back into the project).

From a higher-level perspective, Benkler breaks down the disaggregation described above into three distinct functions<sup>31</sup>. First, there is “an initial utterance of a humanly meaningful statement” (Benkler, 2002). This utterance may be any contribution to any community, be it creative, technical, and regardless of polish or quality. Within Wikipedia, this would include creating a stub article, adding a reference to an article, or even uploading a photograph of a notable landmark to the commons. Second, once the initial utterance has been introduced to the community it must be integrated into a knowledge map of the community members. More simply, this function requires that the utterance is deemed “relevant” or “credible,” with relevance indicating the extent to which the utterance produced aligns with the needs of another community member seeking that information for a single, self-defined purpose and credibility indicating an objective measure of the quality of that utterance. Returning to the latter example of adding a picture to commons in Wikipedia, relevance may be indicated by whether that picture was of the Space Needle but was added to the article on the cone snail<sup>32</sup> (in which case we would argue that the addition of the photograph is *not*, in fact, relevant). And third, that there are adequate means of distributing that utterance. Once again taking the prior example of posting a photo of the Space Needle to the appropriate article, from the perspective of a Wikipedia reader distribution is managed by the platform itself (i.e., the Internet provides a low-cost and effective means of making that utterance widely available). However, from the perspective of those community members *producing* the utterance, distribution may be interpreted as the ability of other community members to maintain accurate awareness of distributed production activities across the community. An example for this latter condition may be the need for a community member to know when a photo of the Space Needle has been posted to the appropriate article before recommending the article be promoted to GA status. In this instance, alternative means of distribution would be required (such as a push mechanism or post to a user talk page, or passive monitoring of watchlists), a process explored more fully in Part III of this dissertation.

Finally, beyond these structural attributes of successful peer production, Benkler states that these communities have two primary characteristics. First, that production is decentralized, indicating that any activity required to maintain project health is within the capacity of the largest portion of the population to complete that activity. And second, that the community relies on social cues and motivations rather than explicit market prices to coordinate member activity. Ultimately, peer production “relies on making an unbounded set of resources available to an unbounded set of agents, who can apply themselves toward an unbounded set of projects” (Benkler, 2002, p. 415).

---

<sup>30</sup> Or, in Benkler’s words, “The number of people who will likely participate in a project is inversely related to the size of the smallest-scale contribution necessary to produce a usable module,” (2002, p. 435).

<sup>31</sup> This was referred to as the two evolutionary phases of online communities in (Viégas, Wattenberg, & McKeon, 2007): creating content (utterance) and quality control (relevance/accreditation).

<sup>32</sup> This specific article was selected at random, <https://en.wikipedia.org/wiki/Conus>, retrieved August, 2015.

### 1.3 COORDINATION THEORY – A BRIEF PRIMER

Combined with the theoretical and structural attributes of commons-based peer production communities, this dissertation utilizes Coordination Theory as an analytical lens. Coordination Theory came out of work completed primarily by Malone & Crowston (1994), extending work previously completed in (Crowston, 1991). In these works, Malone & Crowston present a theory of coordination to describe and identify potential breakdowns between multiple activities and the resources they potentially create or consume. Through identifying a typology of dependencies that can be recognized in processes of information transfer, they suggest common alternative coordination mechanisms that can be adopted to manage those dependencies (Table 1).

But what is coordination? Works cited previously in this section have in part defined coordination as the meta-work that occurs around the production of an artifact. Many works have equated this meta-work with all edits to pages in the various *talk* namespaces (e.g., article talk pages, user talk pages, Wikipedia talk pages, etc). Others have refined that position, instead classifying activity within those talk pages along numerous dimensions, among them coordination. Others, instead, have equated all interactions within a namespace as coordination, suggesting that the entirety of the work comprising those namespaces is coordinative, in effect. Outside these works, the National Science Foundation has previously defined coordination as “the joint efforts of independent communicating actors towards mutually defined goals” as well as “networks of human action and commitments that are enabled by computer and communication technologies” (Malone & Crowston, 1994). Or, more succinctly, Malone & Crowston eventually define coordination as it pertains to their theory as “managing dependencies between activities” (1994). From this perspective coordination can be viewed at the process level – as a means of mediating information transfer between tasks to ensure that each activity’s objectives are met and that the selfsame resources are available for subsequent objectives, ultimately providing a more effective means of both *describing* the paths of information flows through an organization as well as suggesting means of resolving breakdowns in those flows.

One distinguishing feature of Coordination Theory, and a feature that will be recalled and extended upon in future sections of this dissertation, are the organizations observed to create the theory. Crowston (1991) initially observed traditionally top-down organizations in the initial formulation of the coordination typology: an airplane manufacturer, an automobile manufacturer, and a software engineering company. Each of these is a prime example of the markets and firms described by Benkler in previous sections; from (Benkler, 2002):

*Markets and firm-based hierarchies are information processes in the sense that they are means of reducing the uncertainty that agents face in evaluating different courses of action to a level acceptable to the agent as a warranting action. Markets price different levels of effort and resources to signal the relative values of actions in a way that allows individuals to compare actions and calculate the likely actions of other individuals faced with similar pricing of alternative courses of actions. Firms reduce uncertainty by specifying to some individuals what actions to take, thereby reducing the uncertainty of interdependent action to a level acceptable to the agents by delegating to the managers control over enough resources and people by contract and property.*

The distinguishing features remain: in commons-based peer production communities, in which work is primarily self-selected and hierarchy is largely flat, in which resources are unbounded, agents are unbounded, and the capacity to define new projects is unbounded (Benkler, 2002, p. 415), a theory of coordination which seeks to resolve dependencies in organizations with bounded resources, agents, and projects may not have the descriptive power to adequately identify information flow. And accordingly, this theory may not be suitable to suggest future design and research interventions to improve those information flows.

We argue, and much of the sections that follow in this dissertation will cement this argument, for the creation of a theoretical approach to more effectively describe coordination within commons-based peer production communities. Given an understanding of the nature of commons-based peer production communities and those that contribute to them, and given a vocabulary to more effectively describe processes of information flow within those communities (e.g., coordination), we will provide those designers and researchers tools to complete their work, as well as language to both communicate that work to others and to interrogate the objective success of that work.

**Table 1. Malone and Crowston’s common dependencies and example alternative coordination mechanisms for dealing with them. Originally from (Malone & Crowston, 1994).**

<i>Dependency</i>	<i>Alternative coordination mechanism</i>
<i>Shared resources</i>	“First come, first serve,” priority order, budgets, managerial decision
Task assignments	(same as for “Shared resources”)
<i>Producer/Consumer relationships</i>	
Prerequisite constraints	Notification, sequencing, tracking
Transfer	Inventory management (e.g. “just in time,” “Economic order quantity”)
Usability	Standardization, ask users, participatory design
Design for manufacturability	Concurrent engineering
<i>Simultaneity constraints</i>	Scheduling, synchronization
<i>Task/Subtask</i>	Goal selection, task decomposition

## 1.4 CONCLUSION TO PART I

Ultimately, the purpose of this review is to highlight relevant foundational research on Wikipedia and WikiProjects, with particular focus on the theoretical and methodological perspectives that have been used to better understand participation, and more specifically coordination, within distributed teams. Through this understanding I aim to cement the research which follows in this dissertation as an extension of these prior works. The implication of this should be clear. First, the manner through which measurements of activity are achieved and the lens through which they are interpreted can have a dramatic impact on our conception of how distributed teams in commons-based peer production communities interact. As trace data provides researchers with a lower bandwidth means of establishing motivation, it is beneficial to take a multi-faceted approach to interpreting those interactions, to establish observable interactions as proxies for motivation and identification. While rich data may provide utility, it is only its interpretation that provides value. Second, extending from the ability to establish or refute value in distributed interactions comes the ability to more fully describe those interactions. Thick descriptions of a massive, ephemeral community require an understanding of that community that extends beyond observations of individual interactions or aggregations of trace information. An environment that exists as the amalgamation of individual agent’s

motivations and contributions requires a descriptive language capable of encompassing those myriad motivations to adequately recount the varied interactions within it. In this regard, I refer primarily to Benkler's conception of commons-based peer production communities. Third, from the ability to accurately describe the multitude of interactions and information flows comes the ability to theorize the provenance, the meaning, and the intent behind individual agent's activities. This theoretical interpretation provides an abstraction of agent behavior necessary to refine our collective understanding of how work is effectively completed within commons-based peer production communities, as well as the ability to test that theory through both observation and interaction with community members, and through the evaluation of design interventions created to facilitate the work that those agents aim to complete. And finally, with the capacity to refine the theory comes the capacity to refine the intervention; to both fully understand and reflect on the accuracy of the theoretical interpretation of distributed behavior, and to evaluate the efficacy of the intervention intended to ameliorate potential breakdowns gleaned from that theoretical understanding. In the sections that follow, this dissertation will describe the analysis leading to such a theory, the validation of that theory, and the creation of an intervention utilizing that theory to facilitate interactions in one commons-based peer production community.

## Part II – Grounding Coordination: Understanding Interactions in Context

*And there is no living creature, though the whims of eons have put its eyes on  
boggling stalks and clamped it in a carapace, diminished it to a pinpoint and  
given it a taste for mud and stuck it down a well or hid it under a stone, but that  
creature will live on if it can.*

(Robinson, 1980)

*You take delight not in a city's seven or seventy wonders, but in the answer it  
gives to a question of yours.*

(Calvino, 1974)

---

The goal of this section is to enumerate the research that we undertook that led to an understanding of and justification for a theory of coordination for commons-based peer production communities. Following, we introduce two studies completed to highlight and address the vagaries and challenges inherent in analyzing these communities. First, we introduce a study examining the use of the Hot Articles tool – a project-level tool intending to increase group awareness of sites of increased activity on articles within the scope of any given project. Through this study, we introduce the complexity behind the notion of *membership* in commons-based peer production communities as well as the means through which group activity can be interpreted and coordination through task selection can be traced. And second, we introduce a study examining current practice with regard to tool-mediated coordination within Wikipedia, highlighting both the means of identifying those interactions and expanding on the theoretical underpinnings of how that coordination is enacted. Ultimately, this section aims to highlight the process undertaken that laid the foundation for the creation of a theory of coordination for commons-based peer production communities which will be introduced in Part III.

### 2.1 STUDY 1: MANAGING COMPLEXITY: STRATEGIES FOR GROUP AWARENESS AND COORDINATED ACTION IN WIKIPEDIA

In online groups, increasing explicit coordination can increase group cohesion and member productivity. On Wikipedia, WikiProjects employ a variety of explicit coordination mechanisms to motivate and structure member contribution with the goal of creating and improving articles related to particular topics and completing activities related to particular tasks. However, while explicit coordination works well for coordinating article-level actions, coordinating group tasks and tracking progress towards group goals that involve tracking hundreds or thousands of articles over time requires different coordination strategies.

To lower the coordination cost of monitoring and task-routing, WikiProjects centralize coordination activity on WikiProject pages that provide a centralized repository of project tools, tasks and targets, and discussion for explicit group coordination. These tools can facilitate shared awareness of member and non-member editing activity on articles that the project cares about. However, whether these tools are as effective at motivating members as explicit coordination, and whether they elicit the same kind of contributions, has not been studied. In this study, we examine one such tool, Hot Articles, and compare its effect on the editing behavior of WikiProject members with a more common, explicit coordination mechanism: making edit requests on the project talk page.

### **2.1.1 Introduction – Exploring Opportunities for Ambient Information Transfer**

*Joey is a bona fide Football fanatic. In the months leading up to the World Cup he has been pouring his energy into every aspect of Football in Wikipedia. But still, there is more work than Joey can complete. He contributes on the talk pages, but how can Joey maintain awareness of the efforts of the distributed team?*

In many collaborations, understanding where team efforts are focused and maintaining awareness of the other team members' activities is difficult. Shared awareness is realized through information sharing, common knowledge of group and individual activities, and coordination (Dourish & Bellotti, 1992). Such awareness provides a contextual frame in which individual actors make relevant contributions to the activities of other group members.

Wikipedia includes a variety of means for individuals to affiliate and self-identify with others or with specific topics or goals. One of these mechanisms is WikiProjects. For a WikiProject, tools can facilitate shared awareness of member and non-member activity alike. How the awareness provided by these tools might motivate and frame project member activities has not been studied. Like many aspects of Wikipedia WikiProject talk pages are an important place for project members and non-members to communicate with each other. By reviewing the project talk page an editor can get a better sense of what other project members are doing. But, naturally, the project talk page cannot provide a complete picture of all project activity.

Several WikiProjects have adopted the “Hot Articles” tool that maintains a ranked list of articles claimed by the project that are experiencing frequent editing activity. The Hot Articles tool is placed on a WikiProject's main page, providing some awareness of the active content most related to the specific WikiProject. But like explicit requests on talk pages, the Hot Articles tool accounts for only a portion of the potential shared group awareness among participants in the overall system. This study considers how two of these different awareness mechanisms, Hot Articles and WikiProject talk page requests, influence project members' activity. In the sections that follow we outline the related work, introduce our analysis, and then present our study methods, results, and discussion.

### **2.1.2 Two Key Awareness Mechanisms in WikiProjects**

Understanding awareness has been a long running theme in studies of groups and technologies that support group work. Early work on awareness in the workplace (Dourish & Bellotti, 1992) identified key factors that support

awareness and argued for understanding awareness as the context that frames the actions of the individual. Individuals maintain awareness in multiple, varied ways. Both Skeels & Grudin (2009) and Zhao & Rosson (2009) identified means in which activity in social networks may impact ambient awareness of one's own context, defined as the ability to maintain weak ties or to infer status through incremental updates in social networks. This ambient awareness provides a lens to more effectively evaluate tools intended to keep group members apprised of project status within voluntary virtual teams.

Research has been conducted on attempts to automate coordination in Wikipedia through the use of bots (Cosley et al., 2007; Geiger & Halfaker, 2014; Geiger, 2013), to increase directed contribution through project member mediated tools (Zhu, Kraut, & Kittur, 2012), to explore how task lists may facilitate the distributed activities of teams of editors (Krieger, Stark, & Klemmer, 2009), and to better understand or visualize conflict in collaborative spaces (Birnholtz & Ibara, 2012; Kittur et al., 2007; Viégas, Wattenberg, & Dave, 2004). However, there has not been focused research on the potential effects of using automated project-wide tools to increase group awareness by facilitating ambient information exchange.

For this current work, we are focusing on two mechanisms for increasing awareness of activity within the scope of WikiProjects. First, the Hot Articles tool (Figure 6) provides project members with a means of both quickly identifying what article pages associated with the project are being edited most frequently as well as providing a high level view of overall project activity. Conversely, article links in posts to project talk pages typically represent an explicit request for some type of action by those affiliated with the project, thereby increasing community awareness of project needs. In this study, we investigate whether articles listed on the Hot Articles page elicit more editing contributions than articles mentioned in explicit request for participation on the project's talk page.



Figure 6. The Hot Articles page as it would appear on WikiProject pages that subscribe to the tool. Shown above are pages as ranked on WikiProject Feminism (left) and WikiProject Cats (right).

### **2.1.3 Research Hypotheses: Assessing the Impact of the Hot Articles Situational Awareness Mechanism in WikiProjects**

This research has been driven by three primary hypotheses, informed by previous research on awareness and online group collaboration. First, as awareness of project member activity on specific articles is increased through that activity being made explicit via the Hot Articles tool, we suggest that an articles' listing via the Hot Articles tool will result in an increased number of total edits by project members. Hypothesis 1a: *Ambient awareness of project activity mediated by project page updates by the Hot Articles tool will increase the total number of edits by project members to the articles linked.* Similarly, we suggest that in addition to total edits, project member awareness of articles attracting the greatest number of edits in the project will lead to a higher number of unique project members who edit the articles that have been linked to – Hypothesis 1b: *Articles listed by the Hot Articles tool will increase the number of unique project members that edit them.*

Second, as previously stated, WikiProject talk pages are frequently a location where coordination work occurs as individual project members post questions about specific articles, seek opinions about ongoing efforts, or ask for assistance with future updates. Given this accepted and active platform for work coordination, we suggest that an article link included in a project talk page discussion post will direct project member efforts toward that article. And to the extent that the linked article was explicitly posted to the project's talk page, we further suggest that the attention paid to that article will be greater than that seen following an article being linked to by the Hot Articles tool – Hypothesis 2a: *Explicit requests to edit articles on project talk pages will elicit more editing activity by project members than those articles presented in an ambient awareness tool like Hot Articles, measured by total edits by project members to pages linked to on the project talk page.* Similarly, we suggest that explicit requests on the project page will direct a greater amount of attention by unique project members to the articles listed – Hypothesis 2b: *Explicit links to articles from a project's talk page will result in an increase in the number of unique editors who edit the articles linked.*

And finally, as previous research has shown how group identification and common bonds with other group members can increase contributions by members (Kraut & Resnick, 2012; Ren, Kraut, & Kiesler, 2007), we suggest that both of the previous hypotheses will be more significant for project members than non-members – Hypothesis 3: *Both H1a & b and H2a & b will be more pronounced for project members than non-project members.*

### **2.1.4 Research Method: Data Collection and Analysis for the Hot Articles Tool**

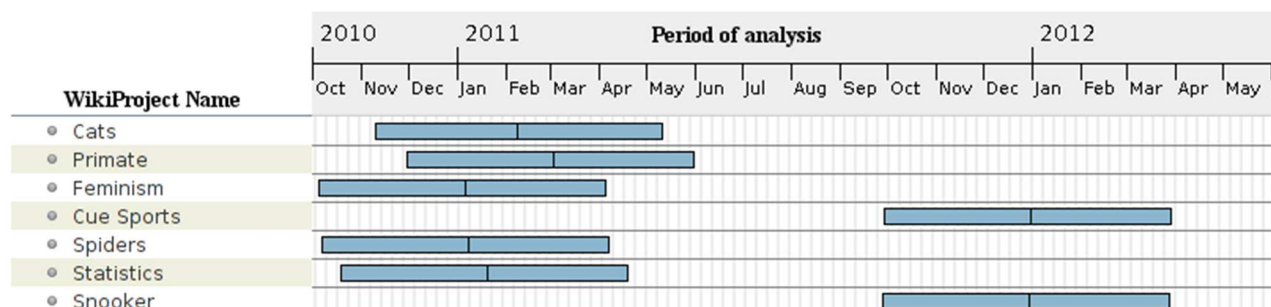
One of the primary goals of our study was to evaluate the effectiveness of an automated tool in coordinating the actions of distributed contributors. To do so, we collected and indexed the complete revision history of every article linked to in a Hot Articles page update for 90 days prior to and 90 days following the activation of the Hot Articles bot for each of the seven projects listed in Figure 7. The 90 day period before and after the Hot Articles bot was activated was chosen to allow us to directly compare within projects the potential effect that such a bot may have on the contributions of project members both before and after the activation date, allowing us to more definitively

distinguish between potential confounds. Specifically, we hoped to be able to show that the top 5 articles edited for a period of time would exhibit unique characteristics in the 90 days prior to the bot being activated (the “pseudo” period), when the true state of which articles are currently seeing lots of activity would not be known to project members, and the 90 days following the bot being activated (the “ha” period), when project members could easily ascertain the most highly active articles by simply viewing the project page, facilitating ambient awareness of project state through passive consumption of the project page contents.

## Data Collection

Project details and all pages owned by each project were collected and locally indexed using the Wikipedia Release Version Tools<sup>33</sup>. This data collection yielded a total of 9,961 pages claimed between the seven WikiProjects under analysis, with the minimum number of pages belonging to WikiProject\_Cue\_sports (818 pages), the maximum belonging to WikiProject\_Statistics (2,379 pages), with a mean of 1,423 and median of 1,158 pages per project.

The data collected for H1a & b includes all revisions to the Hot Articles pages for seven WikiProjects (Figure 7), representing all projects that had subscribed to the Hot Articles bot for at least 90 days before the primary data collection period in September of 2012. Each revision represents an automated update to the Hot Articles page for the given project, and will include the list of the top pages for that project as determined by the edit counts over the last seven days. For each page linked to by the Hot Articles bot, edit counts and unique editors were recorded for the seven days prior to and following that link being posted.



**Figure 7. Gantt chart showing the projects included in the analysis, as well as the time span of the analysis. The black line in the center of each timeline represents the date the Hot Article bot was activated. The period before the bot was active (the “pseudo” period) was included to more effectively determine if changes in article editing patterns was due to the HA bot or other factors.**

Data collected for H2a & b includes all revisions to project talk pages for each of the seven projects under analysis: WikiProject Cue Sports (a project overseeing the varied articles related to sports played with a cue, including 818 article pages); WikiProject Cats (containing articles related to cat breeds, feline disease, and related topics, including 984 article pages); WikiProject Snooker (a cue sport demanding enough attention from editors to deserve its own project, including 1153 article pages); WikiProject Primates (including 1158 article pages related to primates); WikiProject Feminism (including 1455 articles related to feminism, feminist figures and history); WikiProject

<sup>33</sup> Previously, these tools were accessible at <http://toolserver.org/~enwp10/>, *supra*, note 9.

Spiders (including 2014 articles); and WikiProject Statistics (including 2379 articles regarding statistics, including its history and its practice).

For each post to a project talk page, the actual text posted was parsed to find links to pages belonging to that project. Identified links were similarly indexed to determine the total edit counts and total unique editors that modified that page in the seven days prior to and the seven days following that link being posted. All edits and editors recorded were differentiated by project membership status for that point in time, allowing for a more granular analysis of explicit project membership's impact on project related contributions. The seven day period before and after the link was posted for both the Hot Article and Talk page analysis was selected to determine to what extent the posted link had on the editing behaviors of team members. Data was collected directly from Wikipedia Toolserver<sup>34</sup>, allowing full access to the complete revision history to all articles under the purview of the projects under study.

Project members were defined as any editor who added his or her user link to a main project page (e.g., WikiProject Cats) as well as to any other direct descendant of that main project (e.g., WikiProject Cats/Members). If that link was removed at any time between the date that it was added and the current date during the analysis period, membership was assumed to still exist if that user made any edits to the project page during that current analysis period for the current month. All project membership was broken down by month so that granularity of membership could be maintained over the life of each WikiProject. In other words, if User A added her name to WikiProject Cats during month 20 and removed it in month 24, she would be considered a member of that project between the months 20 through 24, inclusive. If she then made edits to the project or project talk pages in months 26 and 28, she would be assumed to be a member for months 20 – 24, 26, and 28, but not for months in which the user link was not present or no edits were made to project pages. If she never added a user link to the project page but still made edits to it, she would not be considered a member of that project. The decision to rely on explicit declarations of membership as opposed to de facto definitions of membership (in other words, editors active within the project space) was made to more directly qualify the types of members expected to be most actively involved in the maintenance and editing of project and related article pages. These distinctions within project membership allowed us a level of detail with regard to analysis of member versus non-member activity that was required for the subtle types of behavioral distinctions we sought to uncover.

## **Analysis**

To test all hypotheses, revisions to the top five most edited articles were recorded for the 90 day period preceding activating the Hot Articles bot to 90 days following the bot's activation, allowing us to analyze editing activity before Hot Articles was activated to provide a baseline for comparison. Edits to the top five articles for each project were calculated on a rolling seven day average (three days in the case of Cue sports), matching the revision information that would have been displayed by the Hot Articles bot on each of the project pages. For each of the articles listed by the Hot Articles bot during the analysis period, we recorded all revisions for the seven days prior to and following that article's listing (again, with three days in the case of Cue sports to match the data recorded and

---

<sup>34</sup> Wikimedia Tool Labs, *supra* note 9.

displayed by the actual Hot Articles page), distinguishing between edits by project members and non project members to inform analysis for H3.

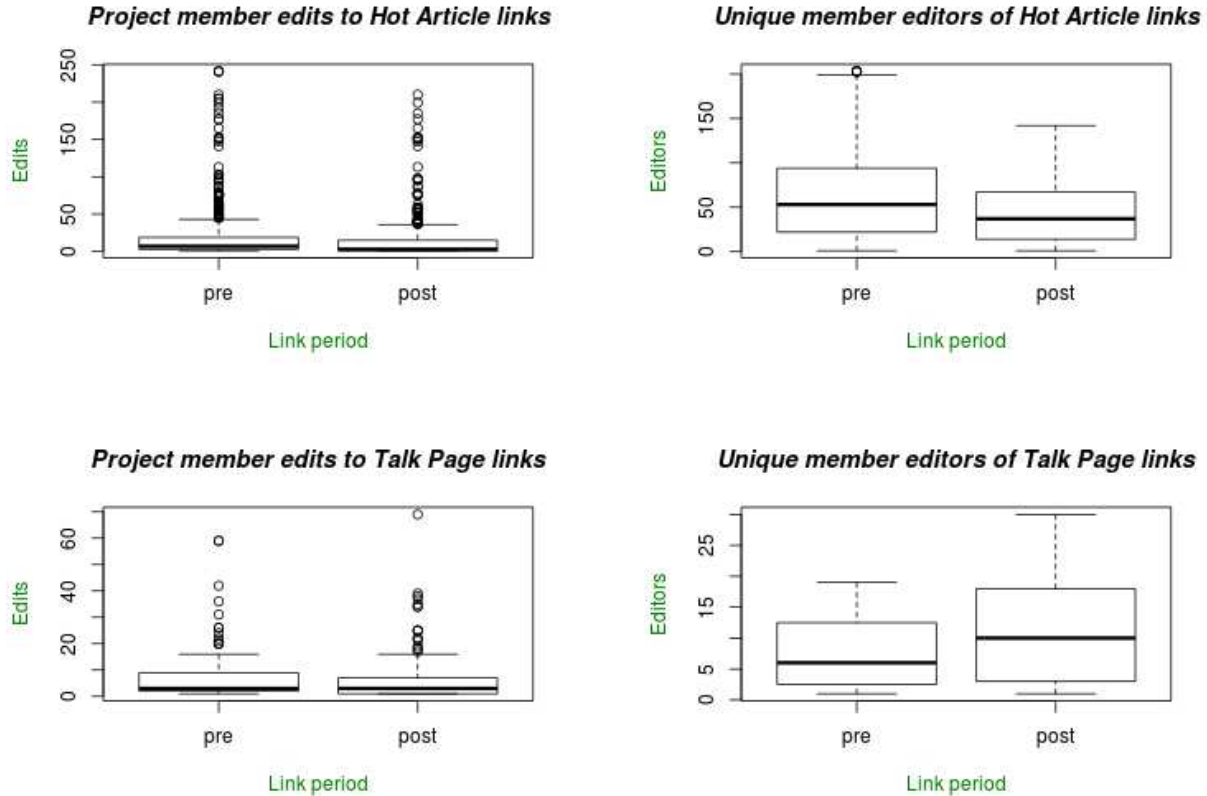
For H1, collected measurements included 1,152 updates to the Hot Articles pages by the Hot Articles bot across all seven projects, yielding a total of 5,760 links to article pages during the analysis period. A total of 43,138 editors edited one of the articles listed in the Hot Articles page in the seven days prior to or following being listed, for a total of 193,523 total revisions recorded during the analysis periods for all projects.

To test H2, we collected all revisions to the project talk pages for each of the above projects during the 90 day analysis period following the launch of the Hot Articles bot. As stated above, a project's talk page is where much of the coordination work typically takes places, with editors discussing issues relevant to the status of the project, including requests to coordinate work activity. For each of these revisions to the project talk pages, the actual text of the revision was analyzed for links to any article pages claimed by that project, a pattern frequently followed when an editor wishes to direct attention to a specific article within the project. Edit counts were then recorded for each of the articles linked to in the project talk page, once again for the seven days prior to and following the link being posted (three days in the case of Cue sports to maintain parity with the earlier analysis), again distinguishing between edits by project members and non project members.

Data collection for H2 yielded 579 revisions to project talk pages across all projects, with 701 editors editing a page linked to in a talk page revision that was a member of the current project in the seven days prior to or following that link being posted, for a total of 2,018 total revisions recorded during the analysis periods for all projects.

## 2.1.5 Results: Analyzing the Impact of Awareness Tools on Levels of Contribution

The results of our analysis, summarized in Figure 8, are split by hypothesis and described below.



**Figure 8.** Means box plots of total edits to Hot Article and Talk Page links and Unique editors to those links. The link period (pre or post) defines the period when project members made the edits – either in the seven days prior to (pre) or following (post) the time when the link was posted.

Hypothesis 1a: *Ambient awareness of project activity through automated updates to the Hot Articles page will increase the total number of edits by project members to the pages linked.*

To test this hypothesis, we compared edits to articles linked by the Hot Articles bot by project members in the seven days (or three, in the case of Cue Sports) immediately preceding and following the link being added to the Hot Articles list for each project. A two group t-test comparing project member edits to Hot Article links in the period before and after a link was posted to the Hot Articles list showed that the Hot Articles link had a marginally significant impact on the total number of edits by project members to the linked pages ( $t=-1.8243$ ,  $df=911.616$ ,  $p=.06843$ ). A two group t-test comparing project member edits to articles listed by the Hot Articles tool in the pseudo condition (the 90 days prior to the date the Hot Articles bot was actually activated) with member edits to articles in the active condition (the 90 days following the activation of the Hot Articles bot) for the seven day period following a page achieving Hot Article status (the post period) was not significant ( $t=-1.5143$ ,  $df=739.583$ ,  $p=$

value=0.1304). Edit counts in the pseudo and active conditions exhibited similar editing patterns, indicating that the links placed on the project page in a Hot Articles list did not have a significant impact on the total edit counts to linked articles by project members.

Hypothesis 1b: *Ambient awareness of project activity through automated updates to the Hot Articles page will increase the number of unique project members that edit the pages linked.*

To test this hypothesis, we compared the total number of unique project member editors in the seven days immediately preceding and following the link being added to the Hot Articles list. A two group t-test comparing unique project members editing linked Hot Articles lists showed that the Hot Articles link was significant, but not in the hypothesized direction ( $t=7.1569$ ,  $df=1128.142$ ,  $p=1.484e-12$ , mean unique editors before the link was posted: 65.15881, mean unique editors after the link was posted: 46.32947), indicating that there are generally a greater number of unique project member editors before an article is posted to the Hot Articles list than after. A two group t-test comparing unique project member editors in the pseudo condition versus the active condition in the seven days following a link being posted was also not significant ( $t=0.1207$ ,  $df=819.84$ ,  $p=0.904$ ).

Prior to an article reaching Hot Article status, it would be expected that its total edit counts and unique editors may shoot up and then regress to the mean shortly after being posted (or, in the case of the pseudo condition, simply reach a peak of edit counts compared to all other articles), and such a pattern would be in line with our hypothesis if the proportion of total edits and unique editors increased between the pseudo condition and the active condition. However, this predicted behavior was not seen for H1a & b. In both cases the means between the active and pseudo conditions in the seven days following a link achieving Hot Article status did not differ significantly. The means of active versus pseudo edits after a link was posted were 14.93968, 18.67519, and the means of active versus pseudo unique editors after a link was posted were 46.32947, 46.02302.

Hypothesis 2a: *Explicit requests to edit articles on project talk pages will elicit more editing activity by project members than ambient awareness, measured by total edits by project members to linked pages.*

To test this hypothesis, we first compared the total number of edits to article links posted in project talk page discussions by project members for the seven days immediately prior to and following the link being added to the talk page. A two group t-test comparing the total edits counts before and after a link was posted to a project talk page showed that the talk page link did not have a significant impact on the number of total edits to the linked page ( $t=.1806$ ,  $df=53.271$ ,  $p=.8573$ ), indicating, somewhat counter-intuitively, that linking to an article page from a project talk page did not have a significant impact on the amount of attention (as measured by an increase in edits) to the page that was linked.

To compare the impact of the Hot Articles bot links with those posted to a project talk page, a two group t-test was done between the number of edits to links posted in the Hot Articles list to the number of edits to links posted to project talk pages, both in the seven days following the link being added. This test proved significant ( $t=3.0068$ ,  $df=99.675$ ,  $p\text{-value}=0.003342$ , mean of edits to Hot Articles links: 14.93968, mean of edits to talk page links: 8.00),

although again not in the hypothesized direction. Links posted in Hot Articles lists received a greater number of edits on average than links posted to the project talk pages.

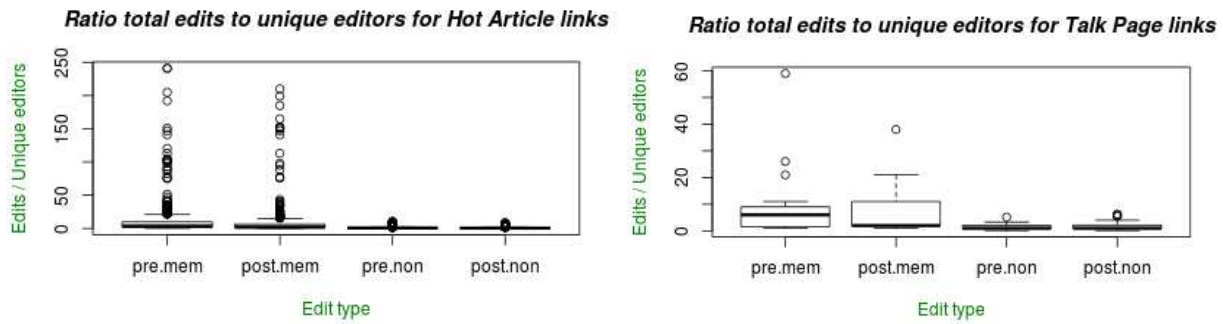
Hypothesis 2b: *Explicit requests to edit articles will result in an increase in the number of unique editors that edit the pages linked.*

To test this hypothesis, we compared the total number of unique project member editors in the seven days prior to and following the link being added to the project talk page. A two group t-test comparing the number of unique project member editors to links posted to project pages in the seven days following the link being added show that links on the project page had a marginally significant impact on the number of unique editors to those pages ( $t=-1.9519$ ,  $df=61.768$ ,  $p=.05549$ ). Links posted to project talk pages may thus potentially increase the number of unique editors to those pages.

To compare the impact on the number of unique project members who edit linked pages between Hot Articles bot links and links placed on project talk pages, a two group t-test was completed. This test compared unique editors between the number of unique editors to Hot Articles links and talk page links, both averaging unique members for the seven days prior to a link being posted. This comparison showed that the difference was significant ( $t=14.9355$ ,  $df=204.336$ ,  $p\text{-value} < 2.2e-16$ , mean of unique editors to Hot Articles links: 46.32947, mean of unique editors to talk page links: 11.24324), although again not in the hypothesized direction. On average, a greater number of unique project members edited pages linked to by the Hot Articles bot than pages linked to on project talk pages in the period following the links being posted.

Hypothesis 3: *The effects for both Hypothesis 1a & b and Hypothesis 2a & b will be more pronounced for project members than non-project members.*

To test this hypothesis we compared the ratio of total edits by unique editors between project members and non-members across both the post and pre periods (Figure 9). For H1, a 2x2 analysis of variance was completed within all projects in which edit ratio of articles linked to by the Hot Articles bot was the dependent variable, and Type (the seven day period prior to or following the article being posted on the Hot Articles page), and Member (whether or not the user editing the article was a project member) were fixed factors. Results were not statistically significant, resulting in ( $f=.276$ ,  $p=.5992$ ). For H2, a 2x2 analysis of variance was completed within all projects in which edit ratio of articles linked to in project talk page posts was the dependent variable, and Type and Member were the dependent variables. The results of this comparison were marginally significant ( $f=2.341$ ,  $p=.127$ ). These results indicate that there is a potential distinction between edits by members or non-members to articles linked to in project discussion pages, but that distinction is not significant in the case of articles linked to by the Hot Articles bot.



**Figure 9. Means box plots of ratio of total edits to unique editors for both edits to articles posted by the Hot Articles bot (left) and edits to articles posted to project talk pages (right). The edit type defines both the period when the edit was made (“pre” or “post”, either in the seven days prior to or following the link being added), and whether the edit was made by a project member (“mem” or “non”).**

## 2.1.6 Discussion

In summary, our findings show that while an article’s listing by the Hot Articles bot had a marginally significant impact on the amount of edits that article received and the number of unique editors contributing to it, that impact was not significantly different from edit counts and unique editors in the period before the Hot Articles bot was active. In other words, the bot may not be accountable for the editing behavior observed (retaining the null for H1a & b). Similarly, we found that edits to articles posted on project talk pages and unique editors of those articles showed only a marginally significant increase (in the case of unique editors), and further that edits and unique editors increased more significantly in the case of articles listed by the Hot Articles bot than direct request, indicating that direct requests for article editing may not be the most effective means of driving further contribution (retain the null for H2a, partial support for H2b). And finally, we found partial support for H3 in that project membership does seem to be a distinguishing feature of edits by project members versus non-members in the case of links posted to project talk pages, but not in the case of links posted by the Hot Articles bot.

## Limitations

There are a number of limitations to the current study that should be addressed in future work. First, there is some debate whether shared awareness tools motivate contributions in a group. Our study assumed that the awareness of article editing activity that would be generated by Hot Article bot lists would motivate project members to pay attention to and edit those articles. This assumption was based on previous research showing that knowledge of group activity and group identification can be an effective motivator to drive future collaboration (Zhu, Kraut, & Kittur, 2012). However, others have offered contrasting theories that may better explain the behavior we observed, which was counter to our expectations. Research lead by Sproull (Sproull, Conley, & Moon, 2005) into prosocial behavior in online settings suggests that increased visibility of potential tasks may cause editors to deem those tasks as less attractive. Such editors may instead prefer to contribute to work in which greater impact can be made due to fewer prior contributions. If potential contributors behave in both the way we expected and in this contrary manner, that it may be that the framing of information in awareness tools either increases or decreases further group

participation. For instance, highlighting articles that have been newly added to a project as opposed to those that have experienced the most recent activity may be a more effective means of coordinating activity among group members.

Second, it is also possible that the nature of articles posted to project talk pages are not strictly comparable to those that are most frequently linked to by the Hot Articles bot. For instance, hot articles currently posted to the WikiProject Cats main page include two examples of animals (Bobcat and the American Lion), one animal known in popular media (Lil Bub), one well known fictional cat (The Cat in the Hat), and one animal related location (Animal shelter). Links posted in this project's talk page at the time of this study include mainly pages that do not currently exist (Brazilian Royal, Ceylon (cat), Karelian Bobtail, etc.), and pages that are potentially less popularly recognized by contributors (articles on the breeds Scottish Fold and Sacred Birman, for instance). This imbalance in the kinds of articles being linked to and their potential connections to knowledge held by the broad editing community may confound direct comparisons to some extent. To address this potential imbalance a more experimental approach may be effective. In such an approach, qualitatively similar articles would be linked to from different points to more effectively highlight distinctive editing patterns between project members and non-members.

Finally, the lack of a highly significant result for H3 calls into question the very notion of how membership can be defined in online contexts. Following an approach used by previous researchers, we defined membership as including all users who had placed their own user link on a project page or membership sub-page, including those who had at one point had a user link and continued to edit project pages or article pages underneath the scope of that project after the user link was removed. If this explicit declaration of membership is unable to differentiate between the collected efforts of contributors to project pages and the articles claimed by those projects, it is possible that a more activity-based operationalization of membership will be required to effectively highlight the nuances between behaviors of explicit members, active participants, and passing contributors. Our primary means of analysis for this study was based on identifying explicit project members, but a more granular approach to membership would provide additional perspective on what types of tools may be most beneficial for different types of project members. It may also provide better insight into what strategies are engaged in by members to coordinate activities at the group level. WikiProjects, after all, are defined as a “group of contributors who want to work together as a team to improve Wikipedia,” not as the task or topic that a group is focused on. Accordingly, we are investigating new techniques to define and determine membership in a more robust manner within and across projects.

## **Implications for Design & Future Work**

Varied explanations are possible for the relative success or failure of an automated awareness tool like Hot Article lists to motivate group activity in a coordinated manner. It may be that coordinated group activity does not follow from such an awareness tool in general. Or, it may be that such a tool can be effective, but not in the timeframes we used to structure our study. Alternatively, it may be that such a tool approach might be more effective for coordinating group activity if it was better tuned to display different or more granular information (e.g., that if Hot Article lists displayed edits distinguished between project members and non-members it may work better, etc).

While initial results show limited promise for this automated tool to enhance ambient awareness in a manner that would better coordinate the actions of distributed contributors, our results must be tempered by the nature of the automated tool chosen to analyze. Namely, it is possible that the increase in editor behavior after an article is listed as a Hot Articles is a phenomenon of the activity that lead the article to be listed in the first place, independent of the article's position on the Hot Articles page. To address this possibility future work should include a qualitative analysis of the articles that saw the most significant Pre to Post changes in editing activity to more accurately ascertain the specific motivations behind those increased contributions, detailing a trace ethnography (Geiger & Ribes, 2011) of the most active editors within our data set.

Insofar as a potential increase in editing activity may be attributed to the article's appearance on a Hot Articles list, a few possibilities exist that may have influenced editing activity. Social proof, for instance, suggests that in ambiguous social situations individuals may take on the activities most prominent around them with the understanding that other contributors have a different or more accurate set of knowledge driving their actions (Social Proof, f. 2015). Thus ambiguity in contributor systems has the capacity to prime individual participants to increase their awareness of available tools intended to motivate them to contribute in the most productive manner, e.g., Hot Articles. Another possibility is that increased awareness of articles that are experiencing the greatest amount of activity could lead to increased contributions to those articles due to a sense of common identity (Ren, Kraut, & Kiesler, 2007). This may be possible because contributions that are in line with the rest of the group that an individual contributor identifies with lead a contributor to identify with that group more intensely, thereby solidifying their position in the group and justifying their continued participation.

Further, while making users' actions visible to other users can increase group activity, it can also lead to conflict. Wikipedia editors sometimes exhibit territorial behaviors around articles they feel a sense of ownership over (Birnholtz & Ibara, 2012; Kittur et al., 2007). Tools that facilitate awareness that an article is being edited can both create new conflict and revive dormant conflicts among editors who are interested in that article. WikiProjects often claim thousands of articles within their area of interest, and WikiProject members may perceive themselves as experts in their topic area. Projects also work to protect the work their members have done on articles from other editors. This suggests that WikiProjects, like Wikipedia editors, may feel territorial about the articles they create and maintain, indicating the potential for automated tools to clarify those boundaries between subject areas and to direct editor motivation towards topics most relevant to that individual.

Finally, while initial data suggests the possibility that automated tools promoting ambient awareness of project activity can successfully coordinate editor action, there is still work that can be done to more fully understand the specific mechanisms through which this coordination occurs. Automated tools intended to increase visibility and motivate contribution should successfully target the population they intend to impact. This is in keeping with Kraut and Resnick's observation that "the selective presentation of tasks can create implicit requests for action, and the better targeted are those implicit requests, the more effective they will be," (Kraut & Resnick, 2012). A tool like Hot Articles that is not ideally effective now could become so if tuned to more accurately correspond to the needs of its intended population. For instance, if a tool similar to the Hot Articles bot were deployed that, beyond simply

displaying project articles ranked by total edit count, also displayed information relevant to the provenance of the editors who made those contributions, it might more effectively coordinate activity than the tool as it is currently deployed. Previous research has shown (Beene et al., 2005; Kraut & Resnick, 2012) that by more accurately directing the message of automated tools towards a specific target population, motivation can be effectively influenced to increase contributor efforts. This suggests the potential for a future comparative study, contrasting the Hot Articles tool with an analogous tool that offered greater granularity of information, such as distinguishing between hot articles edited by project members and non-members, or weighting the list order to reflect edits by the most active contributors to project articles. Yet another comparative study might selectively include non project related articles in a hot articles list to experimentally test prior assumptions.

## **2.2 CONCLUSION TO STUDY 1**

Many WikiProjects use goal-setting mechanisms like collaborations of the week (Zhu, Kraut, & Kittur, 2012) or explicit requests to focus member attention on particular articles. However, the Wikipedia interface does not facilitate easy awareness of activity around WikiProject-claimed articles. Instead, monitoring article activity and tracking improvements to article quality at the project level is often done by hand, or with the assistance of auxiliary automated tools created by community members, such as Article Alerts, External Watchlists, task trackers, and Hot Articles.

Automated tracking tools are often held up as effective mechanisms for assisting with group task-management, change awareness and goal-tracking. However, no previous studies have determined whether the tracking tools used by WikiProjects are as effective at motivating member participation as explicit project-level goal-setting or individual requests for help. In this study we explored how one such automated tool, Hot Articles, has had limited impact on the editing behavior of WikiProject articles by project members. The extent of this impact should be explored further, however. In future work, this research might compare the effectiveness of fully automated bottom-up, emergent applications utilizing ambient awareness to trigger motivation or coordinate activity with the more manual means of coordination frequently at work in online collaborative spaces. Extending from this work, the study that follows introduces a means of identifying such applications and tools and analyzes the impact they may have on coordination within this commons-based peer production community.

## **2.3 STUDY 2: SUPPORTING TOOL-MEDIATED COORDINATION OF VIRTUAL TEAMS IN COMPLEX SYSTEMS**

Support for coordination in online spaces, specifically in peer production systems, has frequently been an afterthought. In the absence of such support, the users of such systems must work to find an emergent order that drives shared project goals and leads to equitable processes. In short, they must rely on the “wisdom of the crowds.”<sup>35</sup> As

---

<sup>35</sup> Recall also the reference to Surowiecki’s “Wisdom of the Crowds” in the email chain from Jimmy Wales from Part I of this dissertation, *supra* note 16.

our study demonstrates, however, the reality is that often the system tools available for coordination, evaluation, and work articulation are not suitable to the task at hand. Our study, first, takes a theoretical approach to understanding how tool-mediated coordination functions within peer production systems. Secondly, we enumerate the methods available to identify automated and semi-automated tools that function within such systems by quantitatively and qualitatively analyzing trace interactions and their utility in Wikipedia over a year-long period. Finally, we identify potential vacuums where new design interventions have the greatest potential for enhancing peer-production systems.

### **2.3.1 Introduction: Situating Tool-Mediation in Current Practice**

In the first week of July, 2014, the Wikipedia user AAlertBot made 237 individual edits to 51 WikiProject pages, notifying project members about articles that may need to be improved, updated, or deleted. In the same week, the user COIBot made 194 revisions notifying Wikipedia users of edits made with potential conflicts of interest. As well, 8,673 unique edits were made by 834 users, rolling back to previous revisions using the AutoWikiBrowser, a semi-automated Wikipedia editing tool. Each of these are examples of bot or tool-mediated edits, the proliferation of which has allowed Wikipedia to grow from a community of like-minded enthusiasts to one of the most visited and recognizable locations on the web today, facilitating a resilient socio-technical system in which individual efforts are amplified and coordination across multiple channels can be delegated to automated and semi-automated tools. In this study, we examine the extent to which such tools impact human activity, and explore how such tools might be designed in the future to better complement human goals.

Group coordination in online spaces has always posed unique challenges. Among them, group members most often are not collocated, frequently with individual team members not even being located in the same time zone. Communication among team members, consequently, can be difficult, sometimes only occurring in asynchronous fashion. Beyond this, knowledge of other group members, their talents, their desires, and their schedules and offline interests are frequently unknown. Policies and social norms are frequently only gleaned through immersion in the group—a process that can only be accomplished through continuous involvement over time, increasing the barrier to entry for new participants who may otherwise have been willing contributors. Rules and governance structures frequently exist, but they too are often impenetrable to new members, ensuring an entrenched minority controls the rules for the majority (Shaw & Hill, 2014). Given these myriad challenges and difficulties, it is indeed remarkable that such projects ever succeed. However, among them, Wikipedia has. And as explored in the sections above, one primary reason for this success is the WikiProject.

To this end, our study offers a quantitative and qualitative analysis of both automated and semi-automated coordination practices in WikiProjects. Our analysis identifies and classifies the myriad ways in which tool-mediated interactions impact coordination work within WikiProjects. In the analysis, tool-mediated interactions refer to the total set of revisions by automated and semi-automated tools, or bots, both within and external to the MediaWiki platform. We classify these tools through a directed content analysis (Hsieh & Shannon, 2005) to determine the spectrum of work activities currently supported by existing tools and to define the types of structured

coordination behavior that could be supported given additional, theory-driven insights. This analysis of existing tool-mediated practices was completed using Benkler's characteristics of commons-based peer production communities (Benkler, 2002) and Malone and Crowston's coordination mechanisms for small group maintenance (Crowston, 2003; Malone & Crowston, 1994).

### **2.3.2 Related Work: Tool-mediation in commons-based peer production communities**

The role of bots and semi-automated tools as coordination mechanisms in peer production systems begins at the intersection of two distinct but related conceptual frameworks: coordination in commons-based peer production communities with respect to WikiProjects and tool mediation, coordination theory, which has been instrumental in defining and modeling the productively synchronized actions of groups. The following sections will expand upon these frameworks with respect to tool mediation, as well as introduce the potential to *theorize tool use* in peer production systems.

#### **Structures and Processes within Commons-Based Peer Production Communities: A Brief Summary**

A commons-based peer production community, according to Benkler (2002), relies on the creativity, intelligence, and efforts of distributed individuals to work towards a common goal. In such communities, production is decentralized. Action within the community is dictated by the desires, talents, and attributes of community members rather than a central organizer, allowing for emergent goals and activities to drive production rather than a single coordinating vision. Beyond this, actors in such communities rely on social cues and motivations rather than explicit market exchanges to coordinate their activity.

Recall from Part I of this dissertation, these peer production communities typically exhibit three types of primary structural attributes (Benkler & Nissenbaum, 2006): first, the objects being acted upon should be modular, allowing for a clean division of potential effort; second, those objects should be relatively granular, with work units that are manageable for the intended population (individual, small group, or large collective) to operate on them effectively; and third, the mechanisms by which changes to the above objects are modified should be capable of being integrated back into the whole community in a simple, low-cost manner.

From the perspective of coordination activity, WikiProjects are modular work elements allowing editors to more effectively locate and define their efforts with regard to their stated membership in projects (Attribute 1). Beneath the project level are tasks, often organized within WikiProjects as the work of Task Forces (Forte, Largo, & Bruckman, 2009), allowing groups to more effectively distribute and organize efforts in a granular fashion (Attribute 2)<sup>36</sup>. Finally, the efforts of these atomic work units can be integrated into the larger encyclopedia and the results of that effort can be recorded on project pages to aid coordination and task tracking for future efforts (Attribute 3), an

---

<sup>36</sup> See also, [http://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Council/Guide/Task\\_forces](http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Council/Guide/Task_forces), for a more detailed description of WikiProject Task Forces

integration and feedback mechanism that is currently not ideally implemented. The potential opportunity associated with this final attribute is expanded upon below.

Given that coordination has been defined as “managing dependencies between activities” (Malone & Crowston, 1994), it is possible to classify and describe such dependencies in any given system to delineate the types of coordination mechanisms that may best ameliorate those dependencies. For example, if one activity is dependent on the successful completion of another activity, coordination mechanisms that facilitate sequential tracking, notifications, or peer review may be useful in driving an overall process to completion. Through the lens of coordination theory, it is possible to define and explore dependencies in project activities and, through doing so, to both identify existing tool-mediated coordination activities within projects and to highlight potential interventions that could be made to improve those processes.

### **Mediating Tools in Peer Production Communities**

In many leading-edge online environments a combination of bots, tools, and scripts substantially shape our interactions with the primary platform. Geiger has labeled this aggregated set of tools as “bespoke code” (Geiger, 2014), referenced in Study 2 above. So beyond the code that drives the core of Wikipedia, beyond the configurations of namespaces, templates, and modules that shape that core, there exists a much greater ecosystem of code that creates and enacts Wikipedia as it is experienced. This includes the bots that patrol for vandalism, grammatical errors, and stylistic omissions, and the scripts that facilitate the ongoing maintenance of the site, the customizable interfaces, and even the bots that send automated greetings for newcomers to the community. Geiger argues (Geiger & Halfaker, 2013b; Geiger, 2014) that these bots and bespoke tools are increasingly ingrained in the systems that people interact within on a daily basis, and that their experiences in those systems are now increasingly shaped, expanded, or constrained by such tools. And, as these tools impact our perception of and means of interactions with a system, they require greater consideration.

Specifically, Geiger and Halfaker (2013b) show how pivotal these tools are to the healthy functioning of the community. They demonstrate that while robust and reliable quality control mechanisms exist within Wikipedia to combat the problems of spam and vandalism, the proper functioning of those tools are required to enable the community to properly function at its current scale. Further, Geiger argues that these tools are no mere force multipliers (2014), simply duplicating human efforts at greater speed and efficiency; instead, through their implementation and manifestation, they are integral to the core experience users have when they visit the site.

These bots, as tools that shape our perceptions of the online spaces we occupy, are worthy of greater attention and exploration. Despite this, it is evident that too little attention has been paid to the potential of tool-mediated systems to facilitate coordination in online spaces. Our study addresses this problem, charting the affordances and dependencies of an existing and active system from both an applied and theoretical perspective. The study explicates the operation and nature of coordination from a detailed analysis of the tools that enable such interactions. By extension, the study then also introduces the potential for new means of coordination within Wikipedia that can be measurably, empirically tested to further our knowledge of how online collaborative spaces could be designed.

### 2.3.3 Data Collection & Analysis

This study makes three contributions to expanding our understanding of coordination in Wikipedia. First, it enumerates the scope of tool-mediated activity within WikiProjects using three distinct methods for data collection. Second, it classifies the nature of those tools through the lens of commons-based peer production communities and coordination theory, expanding existing models to encompass the broader spectrum of functionality provided by those tools. And third, it identifies potential design interventions suggested by the skew in existing versus potential coordination mechanisms fulfilled by existing tools to more effectively address community needs with respect to a modern, socio-technical view of coordination theory.

#### Identifying Automated and Semi-Automated WikiProject Edits

The landscape of Wikipedia contributions has become increasingly complex, with automated, semi-automated, and otherwise tool-mediated edits playing an increasingly important role in the maintenance and ongoing health of the encyclopedia. But, due to the nature of these myriad tools, the largely unstructured ecosystem of bots, user scripts, gadgets, and external applications available to contributors, it is often difficult to identify exactly which edits were tool-mediated<sup>37</sup>. As such, we take a three-pronged approach to identifying these tool-mediated edits with the goal of determining what coordination mechanisms are at work within the scope of WikiProjects.

The sections that follow introduce the method of identifying automated or semi-automated edits as well as present the findings of that method, each in turn. Data for each of the following sections were collected from the Wikimedia Toolserver<sup>38</sup> using custom Python scripts to aggregate and cache all WikiProject revisions in the English Wikipedia for a one year period, starting 05/01/2013. WikiProjects were identified using the category structure present on Wikipedia pages. Included in this data set are all edits to WikiProject pages, sub-pages, or any pages transcluded on either of the above, as well as the corresponding talk pages for each.

#### *Edits by Users with the “bot” Flag*

*Method:* To be an officially recognized bot in Wikipedia, a unique bot account must be created and a request must be made with the Bot Approvals Group where the tasks that the bot aims to complete and the code that will enact those changes can be reviewed, vetted, and ultimately approved or rejected. Once a request is approved, a user account will be added to the “bot” user group, flagging that user as an official legitimate alternative account, capable of editing the encyclopedia in an automated or semi-automated fashion. Edits by these accounts are easily identifiable and provide a simple starting point to identifying the types of activities that automated and semi-automated users complete within the scope of WikiProjects.

*Findings:* In all, bot edits to WikiProject pages comprised 23.8% (507,754 edits) of the total edits to the Wikipedia and Wikipedia talk namespaces, whereas the number of WikiProject pages, subpages, and corresponding talk pages make up only 16.97% (218,611 WikiProject-related pages) of the total pages within those namespaces. This discrepancy, evident in the density plots in Figure 10, is largely the result of a select few bots’ high level of activity

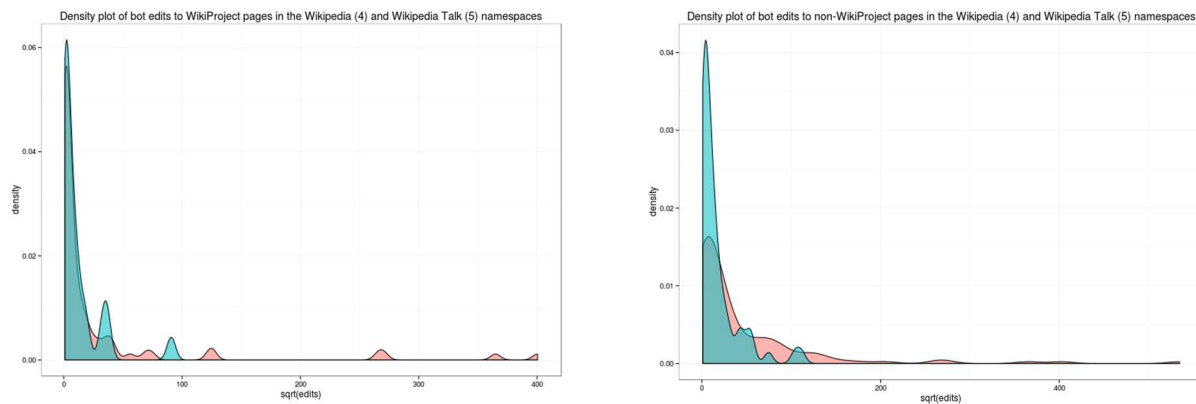
---

<sup>37</sup> Geiger, personal communication, <https://lists.wikimedia.org/mailman/listinfo/wiki-research-l>, “Kill the bots” thread from 5/18/2014

<sup>38</sup> Wikimedia Tool Labs, *supra* note 9.

across multiple projects. The top five bots by edit counts to WikiProject-related pages, for instance, make up 22.32% of the total bot edits, with all remaining edits by bots making up the remaining 1.48% of the total<sup>39</sup>.

Initial analysis of the quantitative data clearly indicates two primary features. First, bot activity is higher among WikiProject-related pages than non-WikiProject pages, notably in the Wikipedia namespace (as opposed to the Wikipedia talk namespace). Second, although there is a greater proportion of bot activity on WikiProject-related pages, the clear majority of that activity is coming from a very few, select bots. The nature of those edits, and the functions they aim to serve within the WikiProject space, will be explored further below.



**Figure 10. Density plot of bot edits to WikiProjects (left) and non-WikiProjects (right). Red represents the Wikipedia namespace (4) and blue represents the Wikipedia Talk namespace (5). X-axis is square root of edit counts.**

### *Non-bot Edits Identified Through Standard Strings in Revision Comments*

*Method:* While it is official policy that bots go through the prior stated review and approval process before operation, there are still accounts that operate bots that have not been approved and therefore do not have the “bot” flag added to the user account, as well as semi-automated tools that facilitate edits on behalf of individual users. To identify these tools, and ultimately to identify the function that they serve within the project space, we additionally looked at all non-bot edits within the project scope, identifying strings that occurred frequently within the revision comments and tags that are added in known tool-mediated edits. For example, tags added in revision comments such as “WP:TW”, “MWT|MWT”, or “AWB|AWB” representing edits completed with Twinkle, Mike’s Wiki Tool, and the Auto Wiki Browser, respectively.

*Findings:* While there are currently no absolute means of identifying tool-mediated edits within Wikipedia<sup>40</sup>, many of the more common tools used within Wikipedia leave identifiable traces through either user names (for instance,

<sup>39</sup> Top five bots by total edit counts to WikiProject visible pages are AAlertBot (233,545 edits), WP 1.0 bot (134,001), COIBot (70,756), JL-Bot (18,723), and Cyberbot I (15,606).

AntiVandalBot, OrphanBot, or PseudoBot) or in the revision comment left with an edit (for instance, for rollbacks, Huggle, Twinkle). The data set comprised of these automated and semi-automated edits was built using identifiers referenced in prior Toolserver documentation<sup>41</sup>, resulting in 346,008 total revisions to WikiProject-related pages during the investigation period. This analysis yielded only 4,874 total tool-mediated edits within the original set of total revisions. Foremost among the tool-mediated edits within the WikiProject space was AutoWikiBrowser, a semi-automated editor that allows users to make multiple edits more quickly and easily than if done by hand, simplifying the process of editing for many more tedious tasks. Second, among the top tool-mediated edits was “undo,” a built-in functionality that enables users to easily, with one click, undo a prior revision on any Wikipedia page. This was followed by Twinkle, a tool that attempts to automate many frequent tasks such as listing a page for speedy deletion, tagging a page for protection, or jumping to the diff of previous revisions of the current page. Combined, these tools represent 96.3% of the tool-mediated revisions identified within the WikiProject space for the period under analysis; however, altogether these tool-mediated edits still only represented 1.41% of the total edits to pages within the WikiProject space. A breakdown of the tools identified is shown in Table 2.

**Table 2. Non-bot edits to WikiProject pages identified by user and revision comment. Not represented are 19 other tools.**

<i>Method of identifying tool</i>	<b>Tool</b>	<b>Edits</b>
<code>UPDATE rev_table SET tool = 'awb' WHERE rev_comment LIKE '%AWB AWB]]%' OR rev_comment LIKE '%AutoWikiBrowser%';</code>	awb	2,121
<code>UPDATE rev_table SET tool = 'undo' WHERE rev_comment LIKE "Undid revision%";</code>	undo	1,786
<code>UPDATE rev_table SET tool = 'twinkle' WHERE rev_comment LIKE '%WP:TW%';</code>	twinkle	790
<code>UPDATE rev_table SET tool = 'huggle' WHERE rev_comment LIKE '%WP:HG%' OR rev_comment LIKE '%WP:HUGGLE%';</code> <code>UPDATE rev_table SET tool = 'huggle' WHERE rev_comment RLIKE "(Message re\\. \\\\[^\]]+\]) (Level [0-9]+ warning re\\. \\\\[^\]]+\])";</code>	huggle	176
<code>UPDATE rev_table SET tool = 'rollback' WHERE rev_comment RLIKE "(Reverted ([0-9]+ )?edits by \\\\[Special:Contributions/[^\\]]+\\\[^\]]+\]) \\\\[User talk:[^\\]]+\\] talk\\]) to last version by .+);"</code>	rollback	1

### ***Non-bot Edits Identified through Successive Revisions***

*Method:* Finally, to identify tool-mediated edits that neither originate from a legitimate bot account nor contain any formulaic revision comment string, we analyzed all edits to pages within the scope of WikiProjects from individual

<sup>40</sup> Geiger, personal communication, see *supra* note 45.

<sup>41</sup> The list documenting how to identify tool-mediated and bot edits through trace data was originally published at [https://wiki.toolserver.org/view/MySQL\\_queries#Automated\\_tool\\_and\\_bot\\_edits](https://wiki.toolserver.org/view/MySQL_queries#Automated_tool_and_bot_edits). However, as the Toolserver, which was hosting this documentation, was decommissioned in June, 2014, I am including the strings used to identify remaining bot and tool-mediated edits in Appendix H.

users that occurred in rapid succession. As Geiger and Halfaker show (2013a), the majority of inter-edit time in edit sessions extends from seconds up to an hour, with the greatest frequency being around one minute. To increase the probability that we are retrieving only tool-mediated edits, therefore, we identified all accounts that made two or more successive edits to any page within the scope of a WikiProject in five seconds or less. Manual inspection of the results of this analysis removed any edits that ultimately did not appear to stem from an automated or semi-automated tool and were simply the result of a user rapidly committing multiple changes to the project.

*Findings:* Using the above dataset of all non-bot revisions to pages within the WikiProject space, including talk pages and pages transcluded within WikiProjects, we reduced the initial set of 346,008 revisions to 25,851 that met the timing criteria. Through manual sorting and selection of frequently occurring strings we identified 26 strings within revision comments that accounted for 97% of the total revisions present in the reduced data set.

Among the top comments identified were “[WP:FWDS|FWDS]” (the trace of a deletion sorting tool), references to “archiving,” “moved page,” “Updating user statistics” (a trace left by an administrator completing batch updates), and “AWB|AWB” (the previously introduced AutoWikiBrowser). Table 3 shows a more detailed view of the top comment strings and the extent to which each was found within the data set, representing the top ~94% of strings by edit count. In all, these concurrent revisions on WikiProject-related pages made up 7.47% of the total number of edits, significantly more than the 1.41% identified through the analysis of non-bot tool-mediated edits shown above.

**Table 3. Non-bot edits to WikiProject pages identified by concurrent edits which occurred <= 5 seconds apart from one another.**

<i>Revision Comment</i>	<b>Edits</b>
<i>[[WP:FWDS FWDS]]</i>	10,873
<i>archive OR archiving (case insensitive)</i>	8,261
<i>moved page</i>	1,582
<i>Updating user statistics</i>	1,475
<i>new section"</i>	805
<i>AWB AWB</i>	644
<i>"" (null comment)</i>	447
<i>WikiProject_Deletion_sorting</i>	250
<i>Other</i>	1,514

### **Classifying Automated and Semi-Automated Activity in the WikiProject Space**

After determining if a bot or tool contributes to a project, we sought to classify how each of them contributes to coordination within the project space. We completed data analysis of automated and semi-automated edits to the project space in two primary ways. First, a mixed methods analysis of bots that have edited WikiProject pages, sub-pages, and pages transcluded within each of them was completed, which aimed to both describe the types of tool-mediated edits that are done to project pages and to show to what extent those edits may have an impact on the coordination of group members, discussed above. Second, we completed a content analysis of these bots, their

functionality, and the boundaries of their utility to WikiProject members with particular focus paid to their potential, realized or not, to coordinate activity of those distributed project members.

Our analysis of WikiProject-contributing bots was completed using a directed content analysis approach (Hsieh & Shannon, 2005) utilizing Crowston’s (2003) coordination mechanisms for small group maintenance as a starting framework for categorizing types of coordination. This directed content analysis allows for a more open inquiry into the specific mechanisms that these bots play in WikiProject coordination and, given their current usage within projects, facilitates the distinctions between coordination mechanisms that may be suggested by coordination theory (Malone & Crowston, 1994) and those that are actually used in practice, highlighting the potential for a novel technological intervention into this space.

An inductive, iterative approach was taken towards identification of automated and semi-automated tool functionality. For each of the 54 bots found to be active on project-related pages for the duration of the study period, we visited the bot user and user talk pages, the bot’s Request for Approval page<sup>42</sup>, and the bot’s contribution history to view both the expressly stated purpose for the bot as well as activity recorded through actual practice. For non-bot tool-mediated edits, we conducted a similar analysis of the documentation page, user manual, or revision history of the related tool to identify the types of coordination mechanism that may be fulfilled by any given tool. Additionally, as many of these automated and semi-automated tools provide multiple types of functionality, where it was identified that more than one coordination mechanism was fulfilled, a tool may be included in more than one category. Classification was completed through activity focused analysis (Malone & Crowston, 1994), whereby the goal is to identify the mechanisms behind an interaction and then search for dependencies that mechanism resolves. For instance, as the primary mechanism behind AAlertBot is to deliver article alerts when pages move in or out of certain workflows (i.e., AfD, PROD, RfC, etc), the dependency that mechanism resolves is the Producer/Consumer relationship, prerequisite constraints, indicating notification, sequence, and tracking. The complete description of all dependencies used to classify tool-mediated activities was adopted from Crowston’s Coordination Theory. If a tool provided functionality not included in the coordination framework, we created a new top-level category to ensure that all tool-mediated interactions were accounted for in our updated classification framework, shown in Table 4, with new top-level categories further defined in the sections that follow.

**Table 4. Prevalence of coordination mechanisms utilized in automated and semi-automated tools in the project space. Percentages given in top-level categories are global, sub-category percentages are relative to each category. Categories observed in practice but not included in the original coordination framework are included under the existing framework.**

<b>Shared Resources</b>	<b>0</b>	<b>0.0%</b>
<i>Task assignments</i>	0	0.0%
<b>Producer/Consumer relationships</b>	<b>19</b>	<b>23.8%</b>

<sup>42</sup> [https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests\\_for\\_approval](https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval)

<i>Prerequisite constraints</i>	10	52.6%
<i>Transfer</i>	0	0.0%
<i>Usability</i>	9	47.4%
<i>Design for manufacturability</i>	0	0.0%
<b>Simultaneity constraints</b>	<b>1</b>	<b>1.3%</b>
<b>Task/Subtask</b>	<b>1</b>	<b>1.3%</b>
<i>Activities not in the coordination framework</i>		
<b>Visibility/Awareness</b>	<b>20</b>	<b>25.0%</b>
<b>Force Multiplier</b>	<b>39</b>	<b>48.8%</b>
Total mechanisms observed	<b>80</b>	<b>100%</b>

### 2.3.4 Discussion

In our analysis, we have confronted two primary research questions. First, how can coordination in online spaces be most effectively mapped and modeled? Second, how might the insight gained from such an analysis inform and direct the design of tools intended to facilitate continued coordination within such peer production systems? While myriad bots, tools, and scripts exist to ameliorate the act of contributing to these online spaces, and WikiProjects in particular, the tools currently in use within these spaces do not fulfill the spectrum of coordinative functions shown by prior research to ensure the optimal functioning of small groups in traditional organizations. Further, we have supported this stance with an empirical analysis of the automated and semi-automated tools used within and beyond the scope of WikiProjects, highlighting the means in which the data was collected, the methods by which that data was interpreted, and the theoretically based justification for the suggestions that resulted—namely, that given the complex nature of the systems many of these tools are created to support, there exists the potential for more focused interventions to directly and positively impact the ability of these online groups to collaborate.

Benkler and Nissenbaum (2006) argue that the mechanisms by which individual efforts are integrated back into the whole should be efficient and low-cost. In WikiProjects, such a mechanism would provide a means by which individual editors would be able to maintain awareness of project activities, project needs, project members, and how they are able to act upon each of these needs. Currently there is no such available tool to facilitate these project-level interactions. The analysis completed in this work aims to fill this vacuum, to foster a peer production community that is resilient, adaptive, and transparent.

## **Discussion of Automated and Semi-Automated Tools**

In (2014), Geiger considers the coordinative power of these automated and semi-automated tools, noting that they are more than mere force multipliers. He argues that the functions these tools fulfill are not simply increasing the amount of work that individuals can complete in online spaces; rather, they fundamentally alter the ways that individuals can complete work in those spaces. We suggest that to serve a coordinative function within the space of an online team, a tool must meet two primary criteria.

First, the tool must in some way manage dependencies of those attempting to participate in group work. This definition extends from the taxonomy of dependencies as outlined by Malone and Crowston (1994), including the producer/consumer relationship (notification of status, transfer of task along dependency chain), simultaneity constraints (effectively manage multiple group members working towards identical ends), and task/subtask dependencies (hierarchy of tasks required for successful completion). Notably, each of these also indicate that any tool-mediation that occurs will require trace evidence in the system in which they are deployed to effectively communicate the nature of the dependency, successfully coordinated or not, to the remaining team members. As we noted earlier, any activity completed in isolation would require no coordinating power or articulation.

Second, the tool must mediate action in a manner that extends beyond being a force multiplier of existing tasks. This requirement stems both from Geiger's (2014) assessment, as well as the nature of the processes that these tools address. From the perspective of coordination theory, there is no transfer of information engendered by these tools that would indicate an increase in coordinative power; consequently, they do not fit within the CT taxonomy. While these types of tools are plentiful, useful, and possibly one of the primary reasons that Wikipedia has continued to thrive, the ability to amplify existing effort by reducing the time to complete arduous tasks does not inherently impact the coordination of remaining group members.

## **Filling the Void: Implications for Theory-Based Design**

While Wikipedia works as a peer production system, our analysis suggests that much more could be done to facilitate complex user interactions within the community of editors. Efforts to quantitatively measure the impact of automated and semi-automated tools have shown that, while these tools are present within WikiProjects, they still support only a small portion of the more arduous work of human edits. Further, our analysis of both the stated and given impact of those tools on project spaces in Wikipedia demonstrates that the primary functionality they provide falls outside what would be considered a coordination mechanism in prior framings of coordination theory.

Given this, we suggest two points for future research. First, while coordination theory may provide a valuable theoretical lens from which to analyze interactions in online spaces, it must be adjusted for explicating interactions in modern commons-based peer production communities. In particular, it needs to be more sensitive to the role of automated or semi-automated tools within such systems. And second, the future tools designed for use within these types of communities should be able to successfully mediate a full range of coordinative actions, thereby removing the greater human burden and facilitating future growth.

The primary evolutionary phases of commons-based peer production communities are disaggregation and aggregation (Benkler, 2002). While many of the automated and semi-automated tools in use within Wikipedia facilitate actions within either of these phases (the force multipliers above), few facilitate organization and interaction among them. For instance, consider an automated tool that provides task recommendations to Wikipedians based on their prior editing activity, or the editing activity of the Wikipedians that they have worked most closely with in the recent past. While prior framings of coordination theory accounts for dependencies such as task assignments, in which work may be completed through managerial decree or organizational requirement, in modern commons-based peer production communities where tasks are self-selected based on personal interests and their completion is motivated by mechanisms outside organizational boundaries, a more salient dependency would require simple awareness. In this case, when the dependency is met, it would indicate that the contributor identified the list of potential tasks and was able to self-select which tasks to complete, or whether to attempt them at all. When such a dependency remains unmet, it is indicative that the contributor was unaware of the work that was being recommended.

This example, in which awareness of a tool-mediated interaction is the primary coordinative function that must be met to maintain an information flow that could lead to task selection and potentially to completing that task (progressing from disaggregation to aggregation), is not accounted for in prior formulations of coordination theory. Accordingly, attempts at theory-driven design utilizing these prior formulations of the theory would be unable to identify the potential failure in the flow of information. We contend that greater attention to these theoretical foundations in the design of future tools will have a positive impact on these online communities as a whole, and the members who populate them.

## **2.4 CONCLUSION TO STUDY 2**

The primary contributions of this study are two-fold. First, while significant research has been completed to understand the role of bots and other automated or semi-automated tools in Wikipedia (see, for instance, Geiger & Halfaker, 2013b; Geiger, 2014), there has been much less focused attention on the use of those bots and tools as a means of structured coordination within WikiProjects. Our study offers an examination of current practice with respect to the theoretical and practical underpinnings of what coordination is, how it is enacted, and what potential there is for future interventions into that coordinative practice. Second, given this more focused understanding of mediating tools for structured collaboration within WikiProjects, we have articulated the potential for design interventions to facilitate project goals and activity, providing a roadmap for developers and researchers to investigate small group functioning in online spaces and suggest more effective and directed interventions for those groups. Ultimately, our analysis shows that coordination theory has the capacity to identify interactions occurring within WikiProjects; however there are far more interactions being completed than the theory has the capacity to describe, suggesting that alternative formulations are required to adequately facilitate theory-driven design.

While the focus of this analysis of coordination in online teams has been on WikiProjects, we expect the knowledge gained from these investigations to be immediately relevant to other similarly distributed teams. What these

investigations aim to accomplish is not only to improve coordination within WikiProjects, but to more clearly define what coordination is within online spaces where humans and automated or semi-automated tools interact, and to expand on our knowledge of its measurement, function, and evaluation.

## **2.5 PART II CONCLUSION**

The prior sections of this dissertation have presented two studies into the nature of coordination in Wikipedia, including the impact of situational awareness on task selection and an examination of current tool-mediated practice within the current landscape of Wikipedia and WikiProject contributors to expand on existing methods of identifying and classifying tool-mediation within Wikipedia. Each of these studies, in turn, lay the foundation for the remaining sections of this dissertation: how to analyze, how to design for, and how to interpret interactions within commons-based peer production communities. In Part III, we extend these foundations through the formulation and introduction of a *theory of coordination for commons-based peer production communities*.

# Part III - “Opportunistic Ad-Hocracies”: Towards a Theory of Coordination for Commons-Based Peer Production Communities

*In physical modeling, the prop is a physical object, while in theoretical modeling, it is usually a prepared description and equation of motion. In some cases, the prop might be a diagram or picture. Just as for novels or paintings, the principles of generation governing the games in which these props function are complex and vary from case to case. In each case, however, the model represents in virtue of prescribing us to imagine things.*

(Toon, 2010)

*The gap between overall structure and underlying components is the symptom of a lack of information: the elements are too numerous, their exact whereabouts are unknown, there exist too many hiatus in their trajectories, and the ways in which they intermingle have not been grasped.*

(Latour, 2010)

---

Extending from the studies in Part II centering on analysis, design, and classification, Part III continues with an in-depth examination of coordination activity within WikiProjects. Through a grounded theory analysis of in-depth, semi-structured interviews, the following sections will first, evaluate the capacity for Coordination Theory (Malone & Crowston, 1994) to identify potential dependencies within commons-based peer production communities, second, introduce a modified theory of coordination to more effectively identify salient human- and tool-mediated interactions within that theory, and third, present both the theoretical and operational output from this modified approach.

## 3.1 INTRODUCTION: REFOCUSING THE PROBLEM

As has been shown in prior sections, coordination can become difficult as the size of the coordinating community grows, and can become further exacerbated due to the distributed and voluntary nature of commons-based peer production communities. The following sections present an objectivist grounded theory of coordination within these

communities intended to both facilitate identification of potential coordination breakdowns as well as introduce means of identifying salient moments of *coordination* across the broad spectrum of activities observed.

Using WikiProjects as a case study, we conducted a series of in-depth, semi-structured interviews with group members of varying experience and motivations to explore types of coordination and information transfer within these voluntary virtual teams. From these interviews, combining a directed content analysis with grounded theory building, we extend prior formulations of coordination theory to more fully describe the types of interactions observed in today's distributed organizations. Our findings suggest that a more robust theory of coordination may be beneficial to both describe potential breakdowns in coordination in modern organizations as well as to suggest alternative means of interaction to resolve those breakdowns, with broader implications for both research on, and theory-driven design of, similar systems. Ultimately, our investigation, grounded in prior formulations of coordination, presents a contemporary theory of coordination for commons-based peer production communities.

In the following sections we first introduce prior work related to the current study. Following, we present the research methods used to explicate our theory from the myriad interactions described by WikiProject members. We then conclude with an analysis, evaluation, and discussion of that theory, highlighting the need for such an approach to better research and design for modern distributed voluntary virtual teams.

## **3.2 RESEARCH QUESTIONS**

This study addresses the question: How can acts of coordination be systematically identified and addressed in commons-based peer production communities? Given prior research into coordination, this primary question covers several subordinate and related research questions:

1. What are the central characteristics of coordination within commons-based peer production communities?
2. How effective is the CT taxonomy in identifying these characteristics?
3. How can this prior theory of coordination be suitably extended to better provide a theoretical tool to identify and resolve potential dependencies within those communities.

In the following sections we first provide a more detailed introduction to the grounded theory method used to classify interactions observed within our data, followed by the process we undertook to approach, interview, analyze, and extend a theory of coordination for CBPP communities.

## **3.3 RESEARCH METHOD**

To build our theory of coordination in CBPP communities, we completed semi-structured interviews with active WikiProject members to explore both current practices as well as the desired mechanisms through which coordination occurs in projects. Interviews were completed with 10 WikiProject members (including one pilot interview), each with a wide variety of experience and involvement with their WikiProjects, spanning from casual content editors to highly active tool developers to long time administrators. Interview participants were recruited

through posts placed on both WikiProject talk pages as well as individual user talk pages, with those individual users identified through active participation in specific projects. In total, eight interview participants were completed over Skype, Google Hangout, or telephone call, with the remaining two being completed via instant message (Table 5 lists details of the interview schedule). Each interview was subsequently transcribed and processed so that each sentence of the transcription was given its own unique identifier to facilitate comparison between coders. In other words, the unit of analysis for all classification was at the level of the individual sentence within each interview transcript.

**Table 5. Details of interview schedule.**

<b>Interview Id</b>	<b>Interview Date</b>	<b>Interview Format</b>
02	01/14/2015	Video
03	01/19/2015	Video
04	01/19/2015	Audio
05	01/21/2015	Instant Message
06	01/27/2015	Audio
07	02/09/2015	Audio
08	02/27/2015	Video
09	03/04/2015	Video
10	04/13/2015	Instant Message

### **3.3.1 A Directed Content Analysis of Coordination**

Our analysis during this phase departs from traditional grounded theory analysis in one notable way: we began with the more deductive directed content analysis rather than the conventional content analysis (Hsieh & Shannon, 2005) more typical of the first phase of grounded data analysis. This choice was made for two primary reasons. First, the nature of our data collection involved semi structured interviews with experienced Wikipedians, with initial questions regarding higher level behaviors and impressions followed by more directed questions about specific types of coordination and collaboration in online teams, with higher level categories informed both by earlier empirical observations and application of existing social science theoretical models to our site of observation: WikiProjects within Wikipedia. The theoretical perspectives selected to inform the lens of our initial observation were commons-based peer production (Benkler, 2002) and coordination theory (Crowston, 1991; Malone & Crowston, 1994). These frameworks were chosen as our initial model of analysis due to their maturity, their relative acceptance and potential in the field of human-centered design and HCI, and due to their descriptive power in interpreting multiple levels of interaction, between humans, systems, and non-human active agents, all within virtual teams. As the nature of our initial line questions extended from an earlier understanding of events from these theoretical perspectives, it follows that analysis of the responses from those questions could extend from them as well. Second, as one of our goals is to test and extend the validity of these theoretical approaches, and the ability to utilize Coordination Theory in

particular to analyze tool-mediated interactions, it follows that the directed content analysis approach will allow us to more explicitly go about that theory testing. Hsieh and Shannon expand on this notion:

*Using existing theory or prior research, researchers begin by identifying key concepts or variables as initial coding categories (Potter & Levine-Donnerstein, 1999). Next, operational definitions for each category are determined using theory...*

*If data are collected primarily through interviews, an open-ended question might be used, followed by targeted questions about the predetermined categories. After an open-ended question, Researcher Y used probes specifically to explore participants' experiences of denial, anger, bargaining, depression, and acceptance.*

As our research questions were informed by the prior analysis of the problem space from the perspective of the above theoretical frameworks, and as the means of data collection included semi-structured interviews intended to more fully understand and explicate the processes by which virtual team members coordinate and collaborate on projects from the understanding of these initial frameworks, we feel that the decision to utilize this directed content analysis approach was justified.

## **Directed Content Analysis Process**

The coding process was completed in two steps. First, all areas in interview transcripts where the interviewee referred to information-seeking or coordination behavior were identified, involving either the users in the system (i.e., Wikipedians), the system itself (including MediaWiki<sup>43</sup>, the underlying software platform which drives Wikipedia), or the tools that help mediate navigation through and interactions within the system (such as the many notions of bespoke code (Geiger, 2014), including the user scripts, gadgets, browser plugins, and the many third party applications operating both online and offline that facilitate both the means of understanding the system and its users as well as the means of interacting with the same). Once these instances of information-seeking or coordination were identified, including all instances of information consumption, information dissemination, or information transfer, the corresponding sentence was highlighted and given an initial code that broadly describes the activity, following guidelines laid out by Charmaz (Charmaz, 2014). For instance, the following transcript sentence was initially coded as “Recruiting group members with tools”:

*03\_0082: And I mean the reason why, the reason why we have that luxury is because we have a bot that's actively recruiting for us, right - We have a bot that's bringing people into our space.*

Second, these instances of information transfer were classified, starting from the vocabulary present in the coordination theory taxonomy (Malone & Crowston, 1994), introducing additional focused codes where necessary to extend that taxonomy to include the broad spectrum of activity possible given the myriad responsibilities and mediating factors acting upon the interviewees<sup>44</sup>. Hsieh and Shannon (2005) substantiate this approach:

---

<sup>43</sup> <https://www.mediawiki.org/wiki/MediaWiki>, retrieved August, 2015.

<sup>44</sup> The first version of the Codesheet utilizing Crowston's Coordination Theory for classification is included in this dissertation as Appendix A.

*The theory or prior research used will guide the discussion of findings. Newly identified categories either offer a contradictory view of the phenomenon or might further refine, extend, and enrich the theory...*

*The main strength of a directed content analysis is that existing theory can be supported and extended. In addition, as research in an area grows, a directed approach makes explicit the reality that researchers are unlikely to be working from the naive perspective that is often viewed as the hallmark of naturalistic designs.*

## Results from the Directed Content Analysis

Evaluation from this initial analysis took two forms. First, we computed the Cohen’s Kappa score to quantitatively identify the level of agreement between coders, accounting for chance agreement (Cohen, 1960). Second, we kept detailed memos about the coding process as suggested by the grounded theory approach to analysis (Charmaz, 2014), ensuring qualitatively that discrepancies identified between the framework and the data could be actively explored.

The initial content analysis was done in an iterative fashion, whereby the codesheet description was updated and refined over numerous passes on the data to ensure the greatest possibility of agreement between coders. Attempting to code our data utilizing only the CT taxonomy along with an additional “Other” category, we achieved an average Cohen’s Kappa of .085. While evaluation of the Kappa statistic is largely subjective and based on the individual field, a frequently cited reference for evaluation (Landis & Koch, 1977) places that score at “Slight agreement” (Table 6). As a result of the poor level of observed fit between Coordination Theory (Malone & Crowston, 1994) and the interactions observed from our data, we progressed from the directed content analysis utilizing Crowston’s Coordination Theory towards building a new grounded theory of coordination for commons-based peer production communities.

**Table 6. General guidelines for evaluating the results of Cohen’s Kappa (Landis & Koch, 1977).**

<b>Kappa Statistic</b>	<b>Strength of Agreement</b>
<0.00	Less than chance agreement
0.00 – 0.20	Slight agreement
0.21 – 0.40	Fair agreement
0.41 – 0.60	Moderate agreement
0.61 – 0.80	Substantial agreement
0.81 – 1.00	Almost perfect agreement

### 3.3.2 Grounded Theory in Practice

Grounded theory analysis typically involves three distinct levels of coding (Corbin & Strauss, 1990; Pace, 2004):

*Open coding:* During the open coding process, interview transcripts were analyzed sentence by sentence to identify instances of information transfer. These initial codes provided a foundation to identify and iteratively refine the focused codes which would eventually resolve into our theory of coordination in CBPP communities.

*Theoretical coding:* During this process, individual concepts and categories were reassembled in terms of their relationships to one another. This included both the temporal aspects of the categories as well as thematic or contextual relationships. Through the process of this organization, a theoretical framework emerged from the data collected which allowed for an organic understanding of the behaviors under observation.

*Selective coding:* The ultimate goal of the prior coding steps was to iteratively, deductively uncover the core explanatory categories that explain the main themes present within the data, each revolving around the sensitizing concept of coordination. Extraneous concepts were trimmed, and data collection continued in a more directed manner throughout the iterative coding process. This process continued until new data no longer expanded the existing categories identified, a milestone known as theoretical saturation.

For our purposes, the above process was completed along two separate lines. First, we completed this grounded analysis of semi-structured interview transcripts, coding interactions identified with both the CT taxonomy in addition to creating and refining a set of open and theoretical codes. And second, as our theoretical codes – emerging directly from the data – diverged from the CT taxonomy as described in the previous section, we progressed to selective coding utilizing our coordination framework for commons-based peer production communities. In other words, initial evaluation was completed with the CT taxonomy, but to effectively create a theory with the descriptive power to identify interactions within our data, it was necessary to codify an alternative theory of coordination. This progression is described in the sections that follow.

## **Towards a theory of coordination for commons-based peer production communities**

Our *theory of coordination for commons-based peer production communities* was also completed iteratively in the form of selective coding of transcript data as well as selective coding the memos that were created throughout the interviews and earlier content analysis process (Charmaz, 2014; Corbin & Strauss, 1990). Through refinement, these selective codes formed the foundation for the codesheet that we utilized to evaluate and validate our theory. Through this evaluation, we identified four primary areas that a theory of coordination for commons-based peer production communities would need to address: identification, awareness, coordination, and production. Each of these top-level categories, as well as specific examples that ultimately informed our coordination framework, are described below.

**Identification:** One of the primary distinguishing features about membership in CBPP communities is the notion that contribution is purely voluntary (Benkler, 2002). Both the type of contribution (i.e., organizational, structural, social, or content-focused) as well as the persistence of the contributor (i.e., a long term editor with a proven focus on improving content within a specific area versus an editor improving ad-hoc articles found through random exploration) are unknown *a priori*, and sensitive to frequent shifts. Contrasted with traditional organizations, in which individuals are often selected for a specific role and then remain in that role indefinitely (i.e., a game designer within a studio rarely switches roles with the 3D animator), modern CBPP communities frequently see contributors

self-identifying with group roles on an as-needed basis, which one interview subject from our study colorfully termed “opportunistic ad-hocracies.” Accordingly, a theory of coordination for CBPP communities would need to account for the capacity of group members to participate in the shifting miasma of needs and availabilities. Specifically, the notion of either explicit or procedural membership within a group provides situational cues to facilitate more structured interactions, allowing group members to more easily discover both who is active within a group at any given time, and what each individual’s primary contributions may be, both issues that are less critical in a more codified organizational hierarchy.

For instance, within our transcript data this would include statements indicating either group or role identification, where identification with a group (of either individual users, a WikiProject, or the community as a whole, depending on the scope of analysis) indicates an active and accepted common identity with regard to the group as a whole (Ren, Kraut, & Kiesler, 2007), and identification with a role indicates an active and accepted role within that group (i.e., task selection). Interviewee’s transcripts give an example of such identification:

*04\_0031: I’ve tended to gravitate toward projects that have personal interest to me... More than anything, personal interest. <Group identification, identifying personally with a given group>*

*03\_0318: But right at the beginning there’s a lot of project management, and it helps to have kind of like more defined leadership roles, so that people are accountable for doing stuff. And also so that you know you can kind of like, you don’t have to spend all your time deliberating about how we’re going to do every little thing. <Role identification, identifying personally with a given role>*

**Awareness:** Related to identification, awareness includes the notion of development within the group, involving interactions that may not leave traces in a system but are still a critical part of maintaining a cognitive model of both the structure of a group, activity within that group, and the health and ongoing maintenance of the myriad bespoke code and tools that allow the group to function. This category is distinct from dependencies in the CT taxonomy as it does not stem from an interaction, but an information flow, related to the capacity of actors to complete new tasks. In Coordination Theory (Crowston, 1991, p.66), knowledge and capability are conflated to simplify the representation of dependencies; however, in CBPP communities where individual roles are far more dynamic than traditional organizations (see Identification above) a more nuanced portrayal of development is required to encompass the broad spectrum of activities identified. This includes the ability to ingrain the social structure of a group, to identify the cultural language of that group to better interact within it, and to more effectively relate group outcomes to potential individual performance (Karau & Williams, 1993). Further, within CBPP communities in which task assignment has largely been supplanted by task selection, in that individual contributions are typically at the whims and availability of their contributors, awareness of task availability is crucial to encouraging work that could be completed.

Again, within our transcript data this would include issues of monitoring, evaluation, and the ability to generate a cognitive model of both the social structure within a group as well as the overall culture of the group, in order to most effectively interact within the group. Interviewee statements corroborate the prevalence of such awareness within the project space:

*02\_0080: More important than being a participant of a WikiProject is understanding Wikipedia community culture. So if somebody's fluent in that, they can go to a WikiProject completely unknown to the people there, start talking, and they'll be heard over somebody new who is trying to learn that particular WikiProject and wants to participate only in that.<Cultural awareness, developing knowledge to successfully communicate using group norms>*

*05\_0031: When I need help, I'll ask someone I already know I can trust or post a request at [WikiProject]. <Social awareness, developing knowledge to successfully identify actors within the group relevant to current needs>*

**Coordination:** By coordination, we refer to the meta-work required to support group processes, pertaining to group members' capacity to maintain the health and ongoing activity within that group. Distinct from the existing CT taxonomy, this category again relates to the needs of group members to interact meaningfully in a community where roles are constantly shifting. More specifically, as new contributors join a group they typically do not have a predetermined role within that group. Given that understanding, it is necessary for documentation, communication, or other group-level processes to exist before that new member can make meaningful contributions towards completing group goals. These coordinative practices then facilitate group awareness, which in turn allow for more meaningful group identification.

Within our data, this would include activities that facilitate both awareness and development of group members. This may include manual or automated updates of project or article activity, improved documentation of project or community practices, or publicly codifying role identification to foster growth or collaboration among the larger group – for instance, to publicly post one's willingness to act as a mentor for new project members. Interview transcript statements indicating a coordination dependency include:

*05\_0083: The project has bots that notify the project (if you look at the main project page) of things like new [project] articles, articles up for FA or GA, etc... <Coordination activity, improving visibility of relevant information through tool-mediated interactions>*

*09\_0062 A: They should have the same template, they should have the same standards, we should have art for these things. <Social coordination, opening debate of group practices in a public forum>*

**Production:** Production pertains to the activities that would more typically fall within the rubric of the CT taxonomy. That is, it refers to the work completed to further group goals. But once again differentiating from CT, these goals are typically met piecemeal by contributors of varying levels of experience or interest, with motives that could include altruism, curiosity, boredom, or malice, and with a governance structure of adequate complexity to attempt to address these differing needs.

Within our data set, this class of dependency designates individual efforts to complete or further group goals. For a topic-oriented project such as WikiProject Cats<sup>45</sup> this may involve adding, editing, or refining content related to cats. Conversely, for a task-oriented project such as WikiProject Copyright Cleanup<sup>46</sup> this dependency may involve identifying, reviewing, or removing potential copyright violations. In other words, regardless of the specific goals of a group, production dependencies exist when those activities are undertaken (i.e., within an open-source project on GitHub, an activity classified as production may be checking in code to fix a bug, adding a new software feature, or refactoring code to improve readability). Example interview transcript statements indicating a production dependency include:

*08\_0009 A: For a long time I just sort of clicked special random, just random article, and you know, Wikipedia I think is probably, just from my experience, probably like 80% stubs or something right, so ya know, I did a lot of stuff like stub categorization, or cleaning up this or that. <Production activity, self-selecting production tasks randomly>*

*08\_0015: ... I guess my work flow would be sort of, I'm clicking around, I see some problem in an article and then I go to the talk page, but then the talk page usually isn't that active, so I look at the WikiProjects that are listed for that, and I go there. Then I leave a message for whoever's there, to say can we get some eyes on this, I don't know what to do. <Social production, engaging group members in collective production task>*

By being sensitive to the dependency categories described above we have a nuanced tool to describe and analyze the types of information transfer found within our data. This focused analytical lens, we found, is required to effectively classify activity in modern commons-based peer production communities, typically lacking traditional hierarchy, populated by volunteers contributing for individual motivations (as opposed to contributing for money), with membership and rates of contribution that are far more ephemeral than traditional organizations (Table 7 lists examples of dependencies identified in interview transcripts, along with brief descriptions to provide context and explanation to justify the dependency).

**Table 7. Sample quotes identifying specific dependencies from interview transcripts. Descriptions under each quote are taken from memos written during the initial interview, transcription, or coding processes (Charmaz, 2014).**

<b>Interview Id</b>	<b>Dependency Identified</b>	<b>Sample Quote Description</b>
03	Role identification	<p><i>03_0111 A: That's not my work for the project.</i></p> <p>There is a distinction between the types of motivations for participating in traditional organizations, frequently centering on money, and the types of motivations for participating in a volunteer capacity as is the case with Commons Based Peer Production communities such as Wikipedia. For the latter, much research has been completed to understand contributor motivations (ie, common bond, common identity, altruism, self fulfillment, etc), but the notion of voluntary</p>

<sup>45</sup> [http://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Cats](http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Cats)

<sup>46</sup> [http://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Copyright\\_Cleanup](http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Copyright_Cleanup)

		participation is not really encapsulated by Crowston’s Coordination Theory taxonomy
05	Cultural awareness, Structural awareness	<p><i>05_0046 A: Sadly, newbies get little help from "the community", they usually learn from hard knocks and all too many long term editors treat them like crap (all wiki not just talking plants here).</i></p> <p>From the perspective of the newbie, the question is not one of motivation but of adequate knowledge of, first, the content area they are attempting to contribute to. And second, knowledge of the community and culture which they are trying to identify with. This second point is frequently glossed over but no less important than the first. Presumably the comment above is less about familiarizing the newbie with the content area (a subject which the newbie is likely already familiar with, given their interest in contributing to its encyclopedia page), instead this mentorship is about making sure the newbie adheres to the cultural and behavioral expectations within the project. This is once again in the realm of development, a notion not accounted for in the CT taxonomy, potentially also indicating a need for structure (ie, codified mentorship roles/tools within projects would presumably resolve much of the above dependency). This mention of a need to ensure that newcomers are adequately nurtured and informed was echoed throughout this interview, with many mentions of both mentorship and support of newcomers, and with many separate but related mentions of selecting membership or tasks based on the project culture (ie, the WikiProject members are just “nice,” an attribute not shared with many other WikiProject members).</p>
03	Cultural awareness, Social coordination	<p><i>03_0105 A: And then sometimes people will call... hosts will call other hosts out for not being nice, which is one of my favorite things to see.</i></p> <p>Related to development, the previous statement was initially coded as “<i>Reprimanding rude hosts.</i>” The implication here was, in order to keep the environment of the project civil (which was a particularly crucial part of this project) it would be necessary to enforce not just the tasks that people did or did not accomplish (typical in most traditional organizations, online or offline), but the manner in which they represent that project... From the individual perspective (which was the context in which the statement above was given) it seems likely that this statement was about solidifying the intrinsic nature of the group, of creating a safe space through the representation of the actors within that space. This also lends credence to the idea that, online, our perception of a space is simply the amalgamation of those other human and non-human actors who have shaped that place before us (for instance, Ribes et al., 2013, or Geiger, 2014).</p>
04	Group identification	<p><i>04_0034 A: The occult stuff is again something that interested me and so I'm into that.</i></p> <p>One of the more common observations of community members in commons-based peer production communities is the flexibility with which they switch from one group to another, from one task to another. Compared to traditional organizations in which tenure in a role is typically strongly correlated with one’s success in that organization (in that, the accountant will typically remain within the</p>

		<p>accounting field, progress within the organization is generally perceived as upward mobility within that profession), identification with any group (let along any role) is frequently fluid, involving an ad hoc selection which may switch at any moment (e.g., for a similar group which allows a closer fit to personal desires given the content under the scope of the group, the context in which the group selects to work, the culture of that group, etc).</p>
10	Group awareness, Social awareness	<p><i>10_0035 A: After trawling through so many articles I think you get an intuitive idea of who's serious about the project and its aims.</i></p> <p>This statement once again hints at the <i>development</i> of community members. In a community where roles are not only not clearly defined but highly fluid, community members are required to build familiarity with the social structures of those in the group to effectively interact with those group members. Without an <i>explicit</i> means of stating one's role within a group (and moreso, without a <i>procedural</i> means of updating that role given the fluid nature of contributions), it is difficult to determine who is the best contact for any given situation without immersion and awareness within the group, <i>both characteristics that are not immediately transferrable to other group members</i>. In this instance, the first part of the statement indicates Group awareness (awareness of content under scope of the group) whereas the second part of the statement indicates Social awareness (awareness of the social structure, or the “movers and shakers” in the group).</p>
07	Social coordination, Production activity	<p><i>07_0125 A: And frequently I don't like to, especially when it looks like it's working in good face, I would like to reach out to them with kind of more personalized touch, sometimes just by moderating the template to make it less like the cookie cutter.</i></p> <p>The distinction between Social coordination and social production is frequently minor. In this case, from the perspective of the user that this interviewee is <i>reaching out to</i>, this communication entails an aspect of Group awareness, in that the recipient of that message is <i>developing knowledge</i> that will help them in future production activities (i.e., how to avoid violating copyright). From the perspective of the interviewee, this simple act involves both <i>Social coordination</i> in that the interviewee is <i>imparting that knowledge</i> that can be recalled in all future interactions by the recipient of the communication with the community, as well as <i>Production activity</i> as the interviewee's communication with the community member fulfills the stated goal of the group – namely, combating copyright violations within Wikipedia. Additionally, this clarifies the distinction between Production activity in topic-focused projects (e.g., WikiProject Cats production entails editing articles on cats) and task-focused projects (e.g., WikiProject Copyright fighters entails removing violating content <i>and</i> communicating that removal to the offending editor).</p>
08	Group awareness	<p><i>08_0076 A: If you look at the quality of medicine based Wikipedia articles, there are a lot of high quality content out there.</i></p> <p>Another trait frequently observed among interviewees was evidence of there <i>general knowledge</i> of the state of a group. This typically would include knowledge of the articles under the scope of a group, but</p>

		would also entail knowledge of the group status (i.e., implicit review of group behavior and success). Looked at from a developmental perspective, this knowledge in turn allows those group members to interact more effectively with regard to necessary coordination and production tasks. For instance, awareness of the group’s faltering ability to track open tasks (with awareness alone not resulting in any trace activity left in the system) could result in the decision to take action to resolve that issue (Role identification) by updating the list of open tasks for a given project (Coordination activity).
06	Cultural awareness	<p><i>06_0084 A: I think when people are level-headed and make good arguments, and argue from the principles of Wikipedia, that’s the best way to involve consensus.</i></p> <p>Similar to observations made in a study led by Rughinis (Rughinis &amp; Matei, 2013), the ability to successfully interact within a group can be negatively affected if the individual interacting is unfamiliar with the group’s norms. In the case of the study referenced above, this was observed as those adopting a “delegated voice” (i.e., talking from the perspective of a Wikipedian, citing policy, seeking consensus, etc). From the perspective of this interviewee, this cultural sensitivity can be seen as an acknowledgement of the “proper” way to achieve consensus within the group.</p>

As stated above, the process taken to construct this theory of coordination within commons-based peer production communities was iterative, involving multiple rounds of coding for each interview transcript, each extending from our initial examination of transcript data from the perspective of the CT taxonomy. The goal of this process was to create a descriptive theory of coordination capable of identifying possible (or past) breaks in an information processing task as well as suggesting prescriptive alternative means of completing that task. This coordination framework is intended to more effectively indicate specific theory-driven design interventions that can facilitate interactions identified within existing processes. The coordination framework is shown in Table 8, with more details on the evaluation and implementation of that theory in the sections that follow.

**Table 8. Coordination framework for analysis of coordination in commons-based peer production communities.**

	Dependency	Common Attributes
Group identification	<b>Group identification</b> (i.e., Self identification with a group)	Project selection, project knowledge, outreach, recruitment.
	<b>Role identification</b> (i.e., Self identification with a task or role)	Task identification, task selection (i.e., personal interest, managerial decree, etc.), role selection (i.e., mentor, tool developer, researcher). Bringing <b>existing knowledge</b> to select Coordination / Production activities.
Group awareness	<b>Group awareness</b> (i.e., Monitoring at the generic level, monitoring at the structural level, group evaluation) <i>Requires NO trace activity in group scope.</i>	Monitoring watchlists, monitoring vandalism. Awareness of content under group scope. Evaluation of group or group members. <b>Developing knowledge</b> to complete Coordination / Production activities.
	<b>Cultural awareness</b> (i.e., Knowledge required to interact productively within group norms, indicating knowledge of the mood, or state of mind of the group or group members).	Knowledge of how interactions should occur (at an interpersonal level), knowledge of shared language, knowledge of group norms, knowledge of policy and practice (not related to production, i.e., “ <i>resistance to change</i> ”), identifying emotional states.
	<b>Social awareness</b> (i.e., Knowledge required to ask the right question of the right project members)	Knowledge of others’ group roles, knowledge to navigate social structure, knowledge of group hierarchy, lurking project communication <b>to understand social structure.</b>
Coordination activity	<b>Coordination activity</b> (i.e., Completing work required to coordinate group activity, tool creation, task creation, structural design, etc.). <i>Requires trace activity in group scope.</i>	Activity required to sustain the group (generic). Activity to improve group organization, maintaining member lists, maintaining task lists, using tools to mediate group activity. <b>Recording knowledge</b> (i.e., documentation), accountability.
	<b>Social coordination</b> (i.e., Communication required to maintain and inform group culture and hierarchy - for instance awareness).	Social activity within group (distinct from group goals), agreement, disagreement, <b>Delivering knowledge</b> related to development (i.e., mentorship), finding collaborators, engaging collaboration.
Production activity	<b>Production activity</b> (i.e., Content creation, completing work to further group goals through modification of non-group artifacts). <i>Requires trace activity in group scope.</i>	Activity to further group goals (generic, distinct from in-group activity), content creation, adding content, cleanup, meta-work, possibly mediated through extrinsic or intrinsic reward.
	<b>Social production</b> (i.e., Communication required to coordinate activities around non-group artifacts).	Communication regarding external artifact (non-group related, although communication can be on group resources), delivering knowledge <i>not</i> related to development.

### *A brief note regarding the split between coordination and production*

With the democratization of the organization comes the need to also democratize the means of coordination within that organization. Since there is typically no formal hierarchy in commons-based peer production communities, every member within those types of communities must be cognizant of and responsible for their own means of contributing to that group. Specifically, where the coordination theory taxonomy is primarily focused on the distribution and use of physical goods or the transfer of information through physical organizations, usually through group members with well-defined roles, this expanded taxonomy is more focused on facilitating the shifting availability and interests of group members, each of whom must self-select roles and tasks, and who are all collectively responsible for the organization and coordination of those tasks.

As the motivations to contribute are distinct, the means to make those contributions are distinct, and the time and talents of each contributor are distinct, this split between in-group and out-group provides the most granular way of defining potential dependencies that could be encountered at each one of these levels, each of which may require alternative coordination mechanisms, and each of those coordination mechanisms could provide differing levels of assistance to each member.

## **3.4 EVALUATING THE THEORY OF COORDINATION FOR COMMONS-BASED PEER PRODUCTION COMMUNITIES**

Evaluation of this theory of coordination took two forms. First, returning to the in depth semi-structured interview transcripts, with each transcript split by sentence and each statement involving an instance of information transfer labeled, we once again coded those interactions in a directed content analysis (Hsieh & Shannon, 2005), this round utilizing our coordination framework as the closed list of categories. Once complete, we quantitatively analyzed these results by computing the Cohen's Kappa statistic (Cohen, 1960) for each round of coding. Second, to highlight the theory's capacity to track dependency trajectories through common tasks, we created multiple scenarios pulled from the interview transcripts, coded each moment of information transfer within those scenarios, and again computed the Cohen's Kappa statistic to evaluate the measure of fit of the theoretical framework with activities pulled from real-world interactions within the system.

Per Haley & Osberg (1989), data conditions to use Kappa are three-fold. First, subjects to be rated are independent of one another. Second, raters score the subjects in an independent fashion. And third, rating categories are mutually exclusive and exhaustive. To ensure the validity of the statistic we recorded, we took steps to isolate individual coder results from the primary research (who dual coded all responses) and split transcripts evenly between group members. Finally, to verify both the mutual exclusivity and exhaustiveness of our coordination framework, we removed all visual or descriptive indications of multi-level hierarchy from the framework (present in the CT taxonomy) and confirmed that dependency coverage of observed interactions was adequately descriptive (i.e., verification that an "Other" category was not being heavily relied upon).

### **3.4.1 Quantitative Evaluation of Coordination Framework**

Similar to the first rounds of evaluation for Crowston's Coordination Theory, this evaluation involved dual coding of each sentence tagged as an instance of information transfer, utilizing the dependencies from our theory of coordination for commons-based peer production communities and their evolving descriptions as our primary codesheet. Between each round of coding, pair-wise disagreements were resolved through adjudication and the codesheet was again iteratively refined to improve descriptions and examples. Ultimately, the average Kappa over the course of all initial coding iterations was .5415, which according to Landis & Koch indicates moderate agreement.

In the social sciences there is no gold standard for evaluation of this statistic; however, the statistic received represents a significant increase over the initial directed content analysis utilizing the CT taxonomy of .085. Through continued refinement of our coordination framework and further codification of the types of interactions indicative of these coordination dependencies, the level of agreement observed in the evaluation of these interactions can still be improved upon before the adjudication process. We suggest that this would be a fruitful avenue for future work.

### **3.4.2 Evaluation of Coordination Framework: Dependency Trajectory Maps of Interview Transcript Scenarios**

As a primary goal of this study was to create a descriptive theory to highlight prescriptive mechanisms to improve group coordination, the final evaluation was completed through classification of group interactions in scenarios. To more effectively map real processes to the potential dependencies that could impede the success of those processes, we used examples from interview transcripts to decompose tasks required to reach some common goals within WikiProjects. Through these scenarios, we illuminate the complexity of common tasks, highlighting the need for continued development, awareness, and collaboration to most effectively contribute to and participate within CBPP communities.

The activities we examined were taken directly from the stories interviewees told. They represented real interactions that acted as the seeds for task decomposition – the deconstruction of the stated activity or goal into its constituent parts, including the context that led up to the task, the development required to complete the task itself, and the means by which the results of each task can be reintegrated into the whole. Dependencies identified within the decomposed task were then utilized as a means of creating concept maps of common activities, including pointing towards the most commonly encountered dependencies and the relationships between them.

The scenarios we constructed attempted to cover a broad spectrum of potential activities and included: 1) Generating audio versions of Wikipedia articles; 2) Dealing with copyright infringers within Wikipedia; 3) Finding collaborators for content production; and 4) Creating tools to facilitate project activity. As each scenario was constructed from the interview transcript of a WikiProject member, unique interactions and trajectories between dependencies, or lack thereof, can provide a novel means of creating a unique fingerprint of each interaction type.

Scenarios were each dual coded by members of the research team, after which all disagreements were resolved through adjudication. Before the adjudication process, the average Kappa statistic for each scenario was .670 showing a continued improvement of coder agreement throughout iterative coding sessions, despite the unique data set (i.e., scenarios versus transcripts). Through theory refinement and continued evolution of our coordination framework and alternative interpretations of individual categories, these initial agreement statistics can be continually improved, and the data analyzed utilizing the framework can venture beyond the WikiProjects that form the case-study for this analysis to include other commons-based peer production communities.

In the process of creating the concept maps, a couple things become clear. *First*, the role of the actor for any given action may frequently shift, despite the fact that the actor has not changed. Within commons-based peer production communities, these types of self-selected role identifications are common, and more importantly, necessary. The nature of the work being completed requires a multitude of roles to ensure that desired actions are effectively planned, coordinated, completed, and evaluated, so the shifting identifications of individual actors is an important component to ensuring that the work is completed and re-integrated into the whole in an efficient manner.

Related to this, during the dependency mapping it will commonly be observed that as an individual's role identification shifts, the lens from which work is being completed will effectively shift as well. For instance, as one individual identifies a broader need within the larger community (such as the need to create a WikiProject to facilitate content creation within a topic area) and acts to create a space to resolve that need, their role will shift from community member to group/project founder. Each of these shifts in self-identified roles precedes the shift in the tasks, actions, and responsibilities commonly associated with that role. Accordingly, to better coordinate this shift in tasks and responsibilities a more explicit mechanism of broadcasting self-identified role may better suit the coordination of group members through codifying the ad-hoc organizational hierarchy. This would have the additional benefits of both increasing group identification and participation through broadcasting a common identity (Farzen et al., 2011) as well as improving the onboarding experience for new group members by specifying more structured potential roles for those group members, as well as more explicitly structuring the means by which potential breakdowns between interdependent actors can be addressed.

*Second*, further distinguishing from the information flow model given in (Crowston, 1991, p.113), the *message* in the observed information flows within commons-based peer production communities is often not explicitly delivered. In these types of communities, where roles are self-selected and contributions are predominantly voluntary, the intended recipients of outcomes from interactions within the community are ambiguous, an ephemeral and constantly shifting group of potential self-identified group members. Accordingly, the sender as well as the recipient of many interactions within that group are frequently the same. That is, as one group member completes tasks related to group goals, it is the mutual and shifting responsibility of both the actant as well as the recipients of that interaction to acknowledge, record, and potentially evaluate the result of the originating interaction. Essentially, this type of broadcast approach to information transfer (asynchronous transfer of information with ambiguous granularity) ensures that each of the individual self-identified roles within the group formation can be fulfilled by

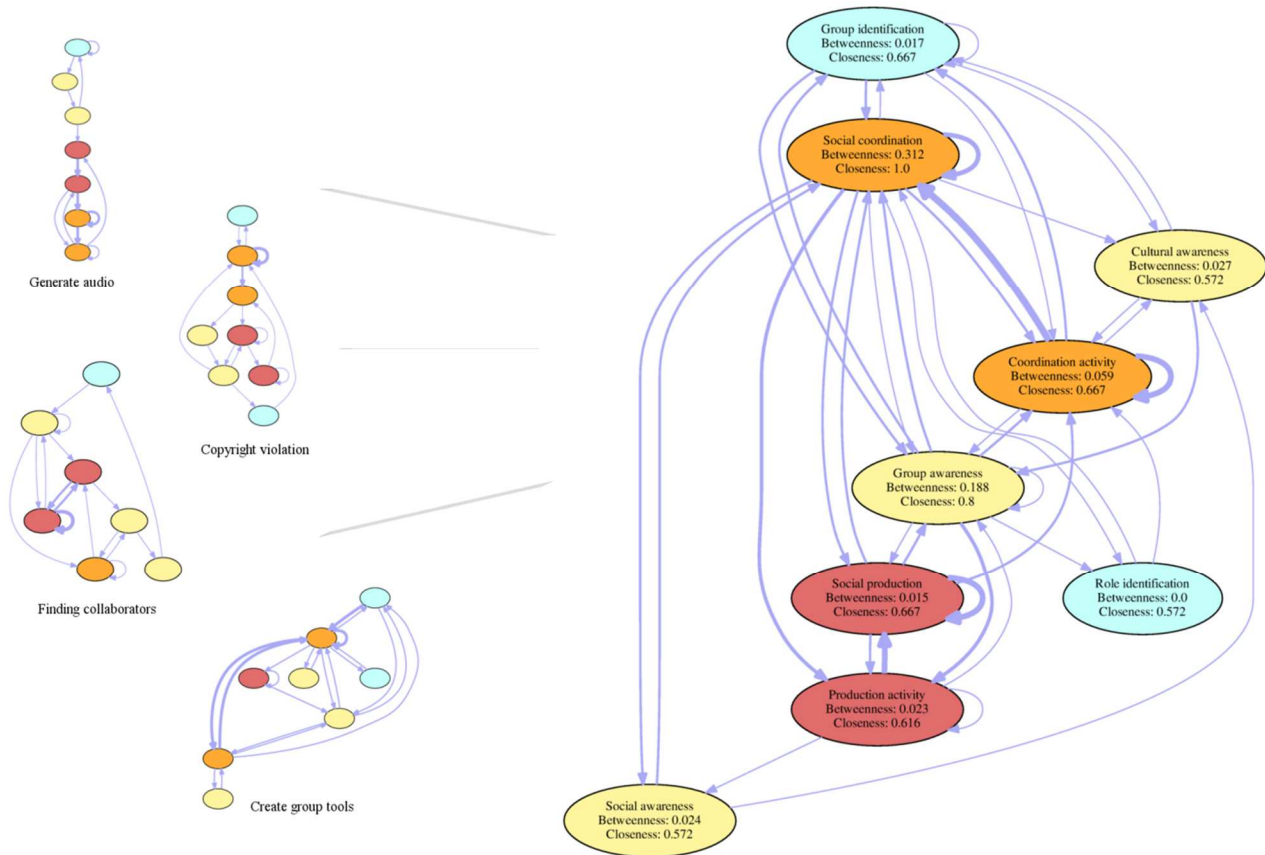
any arbitrary number of actors, including both the human and nonhuman automated and semi-automated tools that may be present within the larger community.

Finally, to illustrate the potential of this theory to highlight not just the individual dependencies identified in myriad interactions, but to additionally more fully enumerate the relationships between those dependencies, Figure 11 shows a concept map of dependencies identified for each individual scenario as well as the combined concept map for all scenarios<sup>47</sup>. Each node in these maps represents a dependency from our coordination framework, with the directed edges between each node representing the dependencies chronologically encountered as we progress through interactions within that scenario. Edge weights are determined by the number of times that specific dependency dyad progression was encountered. This simple diagram provides a number of affordances. First, the ability to quickly identify the most frequent dependency trajectories provides a simple means of targeting future work – namely, focus either on interactions with the most subjective importance or focus on those interactions that are identified most frequently. Second, these dependency trajectories provide a simple means of creating unique dependency concept maps for individual tasks, groups, or communities, providing a novel way of categorizing and grouping common actions within those groups, explored more fully in the future work section below. And finally, a simple social network analysis of the directed network presented in the concept maps provides a simple means of determining centrality measures of each dependency. For instance, in Figure 11 the social coordination node has the highest betweenness of all nodes<sup>48</sup>, possibly indicating that within the scenarios described there is a gatekeeper process, or a bottleneck, worthy of closer inspection.

---

<sup>47</sup> The concept map extends from the grounded theory practice, (Charmaz, 2014, p. 218). The specific implementation of this map is fully reproducible, completed by entering the dependency trajectories in the DOT language and generated with GraphViz.

<sup>48</sup> Betweenness measures the extent to which a node is connected to multiple groups of other nodes who are not otherwise connected to one another (Monge & Contractor, 2003).



**Figure 11. Concept map of dependency trajectories derived from transcript scenarios. Dependencies identified for individual scenarios (left) along with the concept map resulting from combining each of those scenarios (right). The combined map is additionally showing betweenness and closeness scores for each dependency.**

### Walking Through a Dependency Trajectory Concept Map

To more fully understand the utility of the dependency trajectory concept maps presented in Figure 11, consider an example. The scenario presented in the top left was constructed from an interview subject who was a member of WikiProject Spoken Wikipedia, a project whose goal is to create audio versions of every article in the English Wikipedia. The work that this individual was involved with was fairly straightforward – first, sign up to be a member of the group, self-identifying as a group member and audio creator (group and role identification dependencies, identified by the blue node). Following, the individual monitors the project pages to see what standards exist for audio generation and to find high priority articles that might need to be recorded (awareness dependencies, identified by the yellow nodes). Once a high priority article is identified, the individual goes to the appropriate article page and records the needed sections, communicating on that article’s talk page if any questions arise during that process (production dependencies, identified by the red nodes). And once the recording is complete and uploaded, the individual updates the task on the WikiProject page, marking it as complete (coordination dependencies, identified by the orange nodes). This chain between production (recording the audio) and

coordination (updating group tasks) is largely linear; once audio has been created for an article, the individual would simply move on to the next.

Compare this to the scenario in the bottom right of Figure 11, created from the experience of an interview subject responsible for developing tools to facilitate a different WikiProject – the Teahouse, a place where new Wikipedians can ask any questions and get thoughtful, polite, long form responses on any subject. This individual had a very different role within the group, including co-founding the project, identifying needs within that project, creating tools to serve those needs, documenting and deploying those tools, engaging in outreach to ensure the ongoing longevity of the project, and throughout, engaging project members to ensure that activity continually progressed in good faith. Qualitatively, it is immediately apparent that the dependency trajectory observed in the case of the audio generation scenario is far different than the tool development scenario. Where the former follows a relatively linear path through task identification, task selection, task completion, and group update, the latter involves a far more circuitous trajectory. In the latter case, task complexity predicates the possibility that a greater number of potential dependencies will be encountered, each in turn giving rise to the possibility that if that dependency is not met, the process may fail. For example, if the individual within the scenario fails to document the tools they created (coordination dependencies, indicated by the orange nodes) it is possible that those tools will not be usable or will be used improperly by other group members. Or if they fail to engage new Wikipedians to visit the Teahouse through the function of their tools (production dependencies, indicated by the red nodes) it is possible that visitors to the project page will drop off, members will be left silent, and the project as a whole will stagnate. More generally, once the process is depicted in visual form through the dependency trajectory concept map, it is possible to identify *dependency breakdowns*, or instances where a desired interaction does not occur leading to potential failure of that process, providing a novel way to identify areas where processes should be simplified, tools should be developed, or communication should be improved to increase the chance that the process succeeds.

### **3.4.3 Evaluation of Coordination Framework: Dependency Trajectory Maps of WikiProject Interactions**

In addition to the dependency trajectory maps created from the transcript scenarios, we created trajectory maps from the complete history of contributor interactions for select WikiProjects. The WikiProjects selected were among those encountered during the interview process. Care was taken to ensure that the projects selected included a broad spectrum of goals, from the topic-centric projects with the primary goal of improving article content (such as WikiProject Chemistry, primarily dedicated to improving articles containing content on subjects related to chemistry) as well as task-centric projects with the primary goal of improving general processes above the scope of any single project (such as WikiProject Classroom coordination, primarily dedicated to integrating general Wikipedia editing into the classroom curriculum)<sup>49</sup>. In total, 10 projects were selected for data collection and all relevant revisions were collected, utilizing both a local cache of the Wikimedia revision table collected from the Wikimedia Tool Labs servers and the public MediaWiki API.

---

<sup>49</sup> For more detail on the distinctions between topic- and task- oriented WikiProjects, refer to (Morgan, et al, 2014).

The goal for data collection among these projects was to collect all revisions that may indicate a potential dependency within our theory of coordination for commons-based peer production communities. Within WikiProjects, these dependencies were identified using the following criteria:

*Group identification:*

- Indicated by project selection, identified by placing a member link on a project page, sub-page, or any page transcluded on any of those pages, excluding all Talk pages.

*Role identification:*

- For this interpretation of activity, this dependency is implied in the shift between one dependency or task to another (i.e., the *edges* between nodes) and so is excluded from primary collection.

*Group awareness:*

- Indicated by individual awareness of content under the project's scope, identified by all revisions to an article that was linked to by a third party on a project page, sub-page, any page transcluded on either, or corresponding talk page any of the above.

*Cultural awareness:*

- Indicated by shared interactions between project members. This dependency is identified by reciprocal edits between those members, defined as two members interacting on the same User talk pages within the same week. For instance, if project member one and project member two both edit project member one's User talk page, it likely indicates a knowledgeable social interaction which can be interpreted as requiring a level of cultural awareness to successfully participate in the interaction.

*Social awareness:*

- Indicated by project members witnessing social interactions within a project. For example, when an edit is made to a project talk page or a sub-talk page for a project, all subsequent edits by project members indicate implicit awareness of edits by prior project members. To limit the number of dependencies highly active talk pages may trigger, this is limited to prior project member revisions for the prior four weeks.

*Coordination activity:*

- Indicated by project member activity on project pages, identified by all member revisions to project pages, sub-pages, and pages transcluded on any of the above, excluding all related talk pages.

*Social coordination:*

- Similar to coordination activity, social coordination is indicated by member revisions to all project *talk* pages, including sub-talk pages, and the talk pages of any pages transcluded on any of the above, excluding all non-talk pages.

*Production activity:*

- Indicated by all revisions by project members to pages under the scope of that project, excluding all related talk pages.

*Social coordination:*

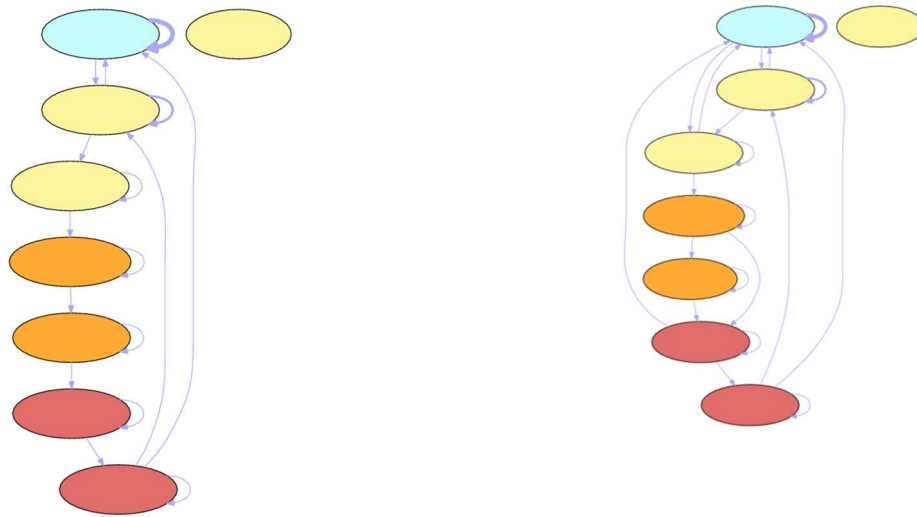
- Indicated by all revisions by project members to *talk* pages of pages under the scope of the project, excluding, of course, all non-talk pages.

Through this operationalization of WikiProject activity, over 4.2 million potential dependencies were identified across the 10 projects data was collected for (Table 9 lists the WikiProjects selected and relevant summary data).

**Table 9. WikiProjects selected for operationalized evaluation of dependency trajectories. The Potential Dependencies Observed includes all dependencies identified over the entire life of the WikiProject. WikiProjects were selected from those encountered in interview transcripts.**

<b>Interviewer Id</b>	<b>WikiProject Name</b>	<b>Topic/Task</b>	<b>Potential Dependencies Observed</b>
02, 06	WikiProject Medicine	Topic	1,500,283
07	WikiProject Missing encyclopedic articles	Task	771,799
01	WikiProject Countering systemic bias	Task	654,274
08	WikiProject Chemistry	Topic	502,164
08	WikiProject Spoken Wikipedia	Task	390,437
01	WikiProject Feminism	Topic	277,758
05	WikiProject Plants	Topic	149,447
07	WikiProject Copyright cleanup	Task	13,245
02	WikiProject Classroom coordination	Task	11,147
04	WikiProject Piracy	Topic	8,197
<b>Total Potential Dependencies Observed</b>			<b>4,278,751</b>

Once data collection was complete, dependency trajectory concept maps were once again created for, first, each WikiProject, second, the top three editors within each project (i.e., the top three members for each project measured by the total number of observed potential dependencies), and third, the combined trajectories of all observed potential dependencies. Figures 12 and 13 illustrate member and project level trajectory maps for two WikiProjects encountered during the interview process.



WikiProject\_Classroom\_coordination

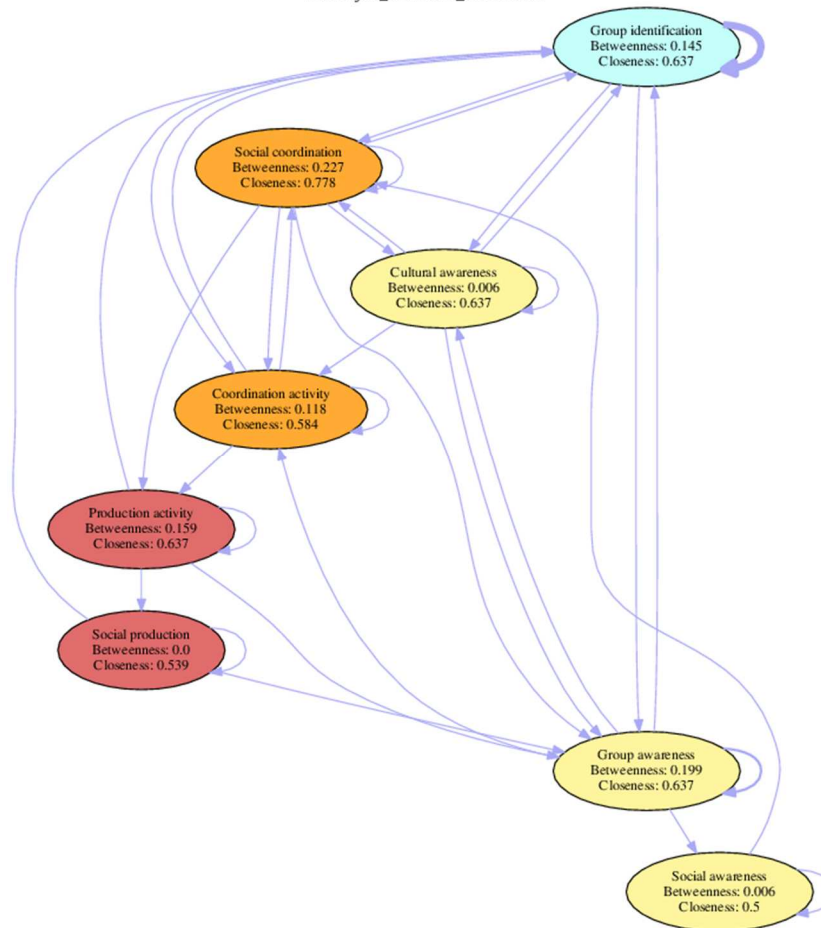
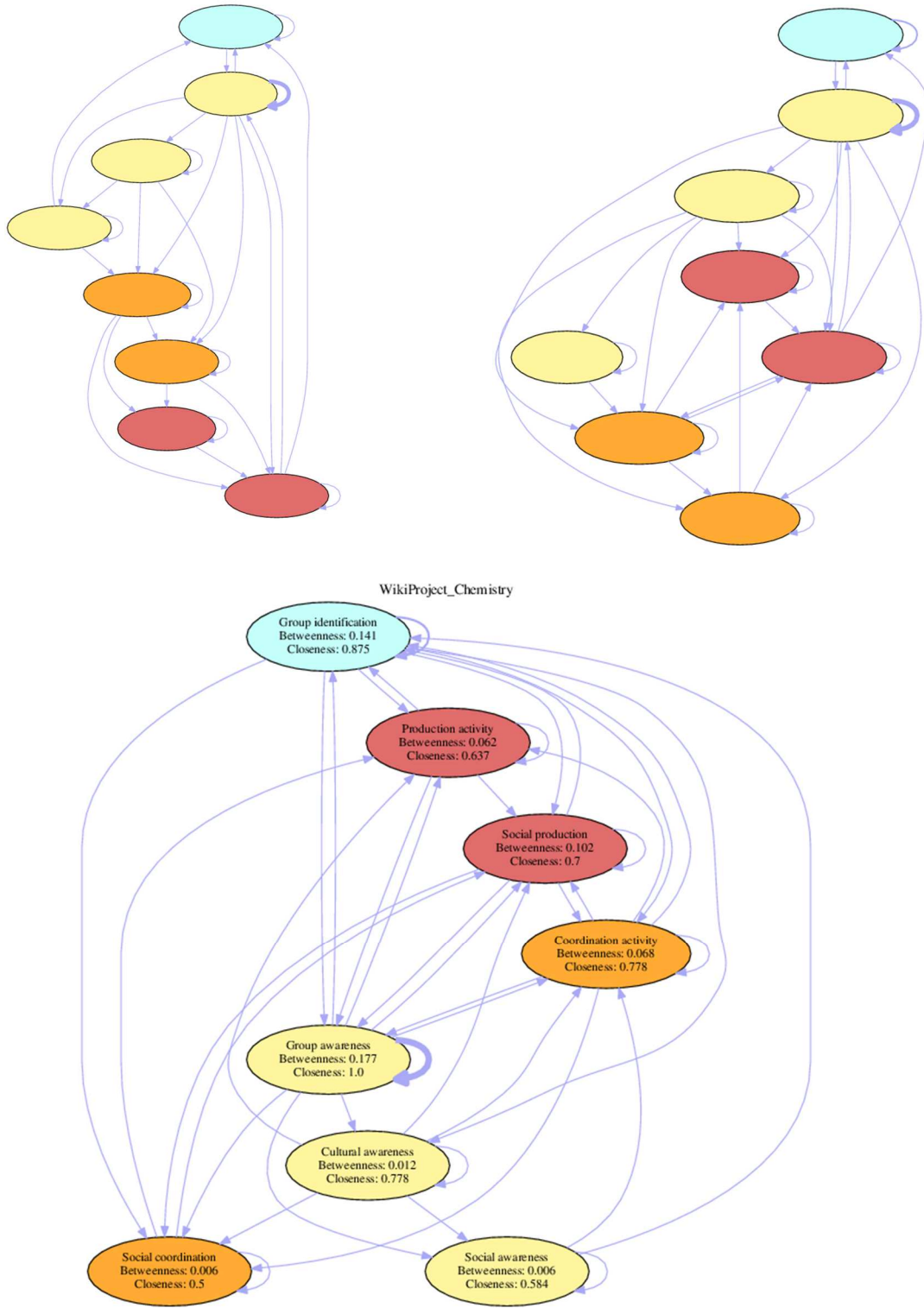


Figure 12. Dependency trajectory maps from WikiProject Classroom coordination member interactions. Top-left and right indicate trajectories for individual users (two project members with the greatest number of observed potential dependencies), while the bottom image includes the combined observed dependencies for the entire WikiProject.



**Figure 13. Dependency trajectory maps from WikiProject Chemistry member interactions. Top-left and right indicate trajectories for individual users, while the bottom image includes the combined observed dependencies for the entire WikiProject.**

While the dependency trajectories included above are merely a starting point, we argue that the flexibility of the means by which each dependency is defined, the variable scale at which each map may be drawn (i.e., at the project member level, at the WikiProject level, at the global level, or combining any of the above), and the granularity of interaction that these representations of activity provide clearly indicate that *coordination in commons-based peer production communities* can be approached with greater precision and nuance than has been achieved in prior work (for instance, in particular compared to prior operationalizations of coordination within Wikipedia as reported in Part I of this dissertation). With this increased descriptive power, we argue, designers would have a more informed foundation for planning future interventions in this types of communities, and researchers would have an increased capacity to validate those designs as well as the information to more effectively expand and refine this theory of coordination for commons-based peer production communities.

### **3.5 FUTURE WORK**

This theoretical coordination framework presents a meaningful step towards providing a uniform way to both identify and potentially resolve dependencies within information transfers in CBPP communities, there remain many directions for future work to progress and for limitations to be addressed. A primary task in this future work will be to extend this theoretical examination to alternative communities. This may include an examination of open source software projects on GitHub, information flows through citizen science initiatives, or even the process of exploring structured deliberation of political discourse using an application like ConsiderIt (Kriplean, Morgan & Freelon, 2011). By continuing to validate and codify the types of information flows described by each dependency, it would be possible to extend our coordination framework to other communities, highlighting potential future breakdowns or analyzing past problems to facilitate and inform more effective design practice.

Second, we have shown above that it is possible to create an algorithmic approach to both identify and map process-centric information flows through publicly available APIs. This approach allows researchers to complete large scale analyses of dependencies within and between groups in commons-based peer production communities, creating a far more nuanced and human-centered view of interactions within those communities. Related to this, it would also enable researchers to create both local and global dependency typologies of these communities, providing a novel way to describe, group, and explore the interactions that are occurring. The work shown above presents a step in this direction. By continuing to refine the operationalization of dependencies within Wikipedia as well as by creating new mappings of these dependencies onto other communities such as GitHub, multi-level analyses of interactions within these disparate communities will provide a novel means of identifying and theorizing the many different types of interactions and identities within each one.

Finally, once dependencies are identified within an organization, there remains the prospect of identifying the alternative coordination mechanisms that have the capacity to resolve or ameliorate those dependencies. Once identified, it would be possible to both examine and experimentally test the capacity of these alternative coordination mechanisms to impact potential dependency failures, creating a roadmap for future organizational design. Beyond the analysis and theory presented in this study, we are continuing to work towards these goals.

### **3.5.1 Justification for this Theory of Coordination for Commons-Based Peer Production Communities (or, why we need more theories of coordination)**

Crowston states, first in (Crowston, 1991, p.29), and then in the text most often cited as the provenance of Coordination Theory (Malone & Crowston, 1994, p.92) that the taxonomy presented should not be considered an exhaustive list of potential dependencies encountered during an examination of an information processing organization, but that it does encompass many of the most frequently encountered dependencies in the organizations he observed. Additionally, while Crowston acknowledges that the actors observed in these interactions may be humans, computer systems, or any other kind of information processor (Crowston, 1991, p.49), he states that the actors he observed in practice were predominantly human. Accordingly, Malone and Crowston's Coordination Theory remains well-suited to identifying information transfer and coordinative processes in more traditional organizations, but without greater emphasis on development and identification their theory has been shown to be inadequate to effectively identify coordination mechanisms necessary to successfully contribute to and interact within a peer production community such as Wikipedia, as demonstrated in the sections above. The site of our observations, WikiProjects within Wikipedia, presents a vastly different landscape of both human and non-human actors, separated from those original observations by a sizable expanse of both time and technology. As such, we expect that through testing this framework in a contemporary, distributed, tool-mediated organization that we will likely find novel behaviors that will allow us to both confirm and expand this new theory of coordination for commons-based peer production communities. The ramifications of this new theory extend far beyond improving our ability to research and design for commons-based peer production communities – as Malone (1987) states, “This perspective [(coordination)] appears particularly promising for understanding how information technology may affect organizational structure because coordination processes are critical in determining organizational structure, and because they are likely to be directly affected by information technology.” That is, by theorizing more powerful tools to describe coordination processes *specifically within* commons-based peer production communities, we are providing new means of analyzing the emergent and sometimes ephemeral organizational structures within those communities.

Finally, by increasing focus on issues of identification, awareness, and development, this theoretical approach is more sensitive to the process of *sensemaking* that must inevitably occur in these types of peer production communities, involving three primary steps (Weick, Sutcliffe, & Obstfeld, 2005). First, is when an individual or individuals translates a flow of circumstances into words and salient categories. This flow of circumstances includes

not just a snapshot of actors in an organization, but also the many and varied “stream[s] of *potential* antecedents and consequences,” (Weick, Sutcliffe, & Obstfeld, 2005). Second, the process of organization is then codified through words and texts. And third, those words and texts are then transmitted through reading, writing, conversation. And third, “reading, writing, conversing, and editing are crucial actions that server as the media through which the invisible hand of institutions shapes conduct,” (Weick, Sutcliffe, & Obstfeld, 2005). Through this lens, this theory of coordination for commons-based peer production communities provides a new way of rendering circumstances (i.e., individual, group, or process level descriptions of coordination dependency trajectories) providing a novel means of attributing how individuals and groups within organizations both make sense of their environments (through *awareness* dependencies, or the first step in sensemaking as described above) as well as how they communicate the product of that awareness and ongoing development (through *coordination* dependencies, involving both the second and third steps in sensemaking as described above). Through the lens of sensemaking, this theoretical approach provides a more nuanced means of analyzing interactions, be they successful or not, and suggesting new ways in which organizational and procedural structures might be modified to better facilitate future interactions.

### **3.6 DISCUSSION**

While our analysis emerges from a desire to better understand how Coordination Theory may more effectively describe coordination mechanisms in CBPP communities, the need to have a more powerful means to describe the commonly seen behaviors within WikiProjects entailed the creation of a modified theory to encompass those more nuanced means of interacting within the system. We anticipate that this theory of coordination for CBPP communities shows a great deal of promise for both the identification and resolution of potential dependency breakdowns within these groups.

Through this analysis, through repeating it within multiple WikiProjects and, ultimately, through multiple corpora and alternative sites, a theory of coordination with the capacity to identify and classify interactions within CBPP communities may emerge, allowing researchers and designers to more effectively analyze, explore, and create new modes of interactions that more effectively facilitate work within these groups. This sentiment is further substantiated by Malone and Crowston (1994), “In baseline analysis, one theory is used as a baseline for comparison to the actual behavior of a system, and deviations from the baseline are then explained with other theories.” Given the unique nature of modern CBPP communities, this study presents such a theory.

### **3.7 CONCLUSION TO PART III**

The previous section explored the provenance, the foundation, the creation, and the initial evaluation of a *theory of coordination for commons-based peer production communities*. While these explorations comprise just the initial steps of a much more ambitious plan to create a more powerful method of theorizing information flow within commons-based peer production communities, we feel that these steps lay the groundwork for a promising new line of research, and in their evaluation provide a concrete proof of concept for this theory. In the section that follows,

we will more fully test this theory of coordination for commons-based peer production communities, utilizing it to drive the research behind and design of a novel prototype application intended to facilitate coordination among online teams – the Virtual Team Explorer.

## Part IV – Creation of the Virtual Team Explorer: Theory applied

*Natural representations are those whose representational powers are constituted independently of the mental states of their users/makers; these would include the number of rings on a tree (representing the age of the tree), the presence of smoke (representing the concomitant presence of fire), and so on. Non-natural representations, by contrast, are produced by human beings for the purpose of communicating something to an audience...*

(Callender & Cohen, 2006)

*Work stops at sunset. Darkness falls over the building site. The sky is filled with stars. "There is the blueprint," they say.*

(Calvino, 1972)

---

In Part III of this dissertation I described the process of building a *theory of coordination for commons-based peer production communities*. This theory was grounded in the experiences of community members and evaluated utilizing interview transcript data, scenarios constructed from those data, and a collection of over 4.2 million observed dependencies within a set of WikiProjects. The culmination of this theory is a novel means of analyzing and interpreting interactions within peer-production communities, providing researchers with a more precise methodology to examine interactions at multiple scales, including the individual, the group, or the global community. Further, the outcomes of these analyses provide designers with specific, actionable data and insights to build more effective interventions within those communities.

Part IV of this dissertation presents an exploration of such an intervention. In this section of the dissertation, I utilize the theory of coordination for commons-based peer production communities to inspire and inform the design of a tool intended to facilitate coordination within those distributed communities. Using this theoretical understanding of individual, group, and community-level interactions within Wikipedia, and building on the coordination needs of online communities identified in prior research, I construct a set of design requirements for a new tool-based approach to coordination. The outcome of this theory-driven design intervention is the *Virtual Team Explorer*, an application intended to facilitate group work within WikiProjects in Wikipedia. As I demonstrate in the following sections of this dissertation, the Virtual Team Explorer is designed to coordinate interactions both within and between these projects, and to enable individuals to more effectively contribute to the larger Wikipedia community.

## 4.1 INTRODUCTION FOR THE INTERVENTION

Vir-tu-al Team Ex-plor-er - /ˈvərCH(əw)əl tēm ikˈsplōrər/ – *noun* – A tool intended to facilitate coordination among distributed teams through multi-modal management of information structures.

Alternatively: VTE

The popularity and longevity of Wikipedia stems not from the litany of clicks from the curious and inquisitive, however, but from the consistent efforts of the contributors who create and improve the content therein. If there was nothing worthwhile to look at, after all, who would come to visit?

In the sections that follow, I introduce a novel architecture designed to increase the potential for both individual editors and collaborative groups to participate with the Wikipedia community in a manner more conducive to distributed content creation. Further, through the creation and publication of this architecture, I hope to engage both researchers and practitioners in a more active discussion regarding the potential for tool-mediated interventions to improve coordination – the latter so that their concerns can be heard and addressed, and the former so that the means and methods of coordination work within WikiProjects in Wikipedia can be better understood to inform and refine the theoretical foundations utilized to design the suggested intervention. More specifically, while Part III of this dissertation aimed to create and evaluate a theory of coordination for commons-based peer production communities, Part IV aims to explicitly utilize that theory in an effort to test its capacity to drive future development efforts.

Part IV of this dissertation is broken into five sections. First, I revisit relevant research into coordination within online groups, and Wikipedia in particular, discussing salient design implications that come from such work and more specifically situate that research with regards to our theory of coordination for commons-based peer production communities. Second, I introduce the architecture created to support such a design, including the tools required to replicate such architecture. Third, I introduce the Virtual Team Explorer (VTE), an application developed to facilitate coordination within WikiProjects, taking into consideration Wikipedians' varied needs and desires explicated in prior literature as well as our theory of coordination. Fourth, I present a discussion of the drawbacks and merits of such a tool including the results of a survey of all interview subjects introduced from Part III. Fifth, and finally, I conclude with a discussion about future work that could be completed in an effort to continue to improve design interventions for peer production communities as well as our capacity to analyze and interpret interactions within those communities through further theory development and evaluation.

## 4.2 RELATED WORK: CONSOLIDATING INTERPRETATIONS OF COORDINATION

The design prototype and underlying architecture presented in the following sections extends from two sources. First, it extends from a review of prior research on tools to facilitate coordination in peer production communities. This review builds on those works cited in Part I of this dissertation, expanding the focus from theoretical and operationalized interpretations of coordination to include research-oriented design interventions within those communities. And second, the design prototype extends from our theory of coordination for commons-based peer

production communities introduced in Part III, and is intended to provide a concrete example of how that theory can be utilized to provide explicit, actionable data to inform tool development as well as to improve the capacity for developers to evaluate the merits and relative success of that development. This section will first focus on the former, with the aim to revisit this prior design research on Wikipedia and WikiProjects through the lens of our theory of coordination in commons-based peer production communities, following.

#### **4.2.1 Situating the Virtual Team Explorer: Review and Recap**

Recall briefly that Malone and Crowston (Malone & Crowston, 1994) defined coordination as “managing dependencies between activities.” These “activities” include any instance of information transfer between organizations, where organizations can refer to “a collection of intercommunicating actors, where in principal the actors can be human beings, computer systems or any other kind of information processor” (Crowston, 1991). At this level, coordination is then interpreted not only as the means of producing a product, but the context of the provenance of that product. In other words, the study of coordination in modern commons-based peer production communities must remain sensitive to the typically distributed, voluntary nature of those communities, as should the design interventions that result from such studies. By situating the theory of coordination for commons-based peer production communities in an examination of actual lived experiences within one of those communities, we have aimed to ensure that the theoretical perspective is properly situated to more effectively facilitate description and analysis of those communities.

This theory-focused view provides a useful lens for evaluating the potential for tool-mediated interventions to impact the interconnectedness of actors within such systems, specifically with regards to how coordination may, or may not, be impacted by such interventions. For example, E. Gilbert (2012) created a tool to resolve a triadic awareness problem in Twitter whereby friends of friends receive duplicate information due to a lack of translucence in prior interactions with the system. This tool affords users a novel (and entertaining) means of shaping their communications, but the nature of that content and the context in which it is delivered remains unchanged. McDonald et al. (McDonald, Gokhman, & Zachry, 2012) explored how architectures for translucence are able to explicate relationships among editors in Wikipedia that would otherwise be time-consuming or impossible to unearth. These relationships provide a means of identifying valued contributions or potential collaborators. However, without direct intervention to impact the impediments to coordination that exist within that system, the intervention does little to resolve underlying issues of identifying suitable collaborators or tasks. Similarly, visualizations such as WikiDashboard (Chi & Suh, 2008), the Hot Articles tool (Gilbert et al., 2013), or tools such as SuggestBot (Cosley et al., 2007) provide a novel means of consuming information about artifacts. Individually, though, these visualizations and tools do little to resolve the “dependencies between activities” that impact coordination at a group level. In other words, from the perspective of our theory of coordination for commons-based peer production communities, each of these tools typically resolves a single dependency. But from the context of the process that generally needs to be solved, utilizing the same vocabulary from our theory of coordination, these are but just one dependency among many that may break down within any of those processes that community members

are attempting to complete (this example will be further expanded upon and explained in the next section of this dissertation).

## Articles you might like to edit, from SuggestBot [\[edit\]](#)

---

SuggestBot predicts that you will enjoy editing some of these articles. Have fun!

### Add sources

- [Computer](#)
- [Design flow \(EDA\)](#)
- [Radio-frequency engineering](#)
- [PARC \(company\)](#)
- [Small-world experiment](#)
- [IBM Common User Access](#)

### Cleanup

- [No Innovation Without Representation](#)
- [Neuroinformatics](#)
- [Mobile phone overuse](#)

### Expand

- [Portable media player](#)
- [Vehicular communication systems](#)
- [Smartwatch](#)

### Unencyclopaedic

- [Web usability](#)
- [Computational creativity](#)
- [Virtual reality](#)

### Wikify

- [Multimodal interaction](#)
- [Cellular neural network](#)
- [Optical head-mounted display](#)

### Orphan

- [Virtual world framework](#)
- [Agile usability engineering](#)
- [Lateral computing](#)

### Merge

- [Information and communication technologies for development](#)
- [GNOME](#)
- [Software verification and validation](#)

### Stub

- [Georgia Tech Information Security Center](#)
- [SIGCHI](#)
- [John M. Carroll \(information scientist\)](#)
- [Algorithm design](#)
- [Hudl](#)
- [ACM Computing Classification System](#)

**Figure 14. Recommendations for tasks to complete created by the SuggestBot tool. Articles shown are related to the Human Computer Interaction Articles category within Wikipedia.**

Another study led by Forte (Forte et al., 2012) confronts issues of coordination more broadly, utilizing McGrath's typology of group modes and functions (McGrath, 1991) to argue for the need to understand coordination not merely as a means of improving production. Instead, an expanded view of coordination is advanced, including notions of group well-being and member support in research and design practice. In this work, Forte references Beschastnikh et al. (2008) and Kriplean et al. (2007), which both examine the impact of policy citations in Wikipedia article talk pages, as examples of work that illustrates broader social impact of coordinated interactions. Specifically, policy citations on Wikipedia talk pages are not only a means of grounding discourse within the cultural norms of Wikipedia, but also a means for editors to establish social roles, to make power plays, and to interact in a way that promotes a common group identity through that shared knowledge. Kriplean further explores the design implications of these findings, suggesting that while a tool-mediated solution could be introduced that

would facilitate a more streamlined or simple exchange among users, it also has the problematic potential to minimize the social exposure among new and experienced members, creating a scenario where enculturation and development could be traded for ease of use. Lacking this socialization, new members are at a disadvantage as they attempt to internalize the means of communication adopted by the larger community.

In particular, this last finding by Kriplean et al. appears quite salient in the context of this dissertation. While coordination has been a central focus in a fair amount of research on Wikipedia and WikiProjects (refer to Part I for the original review), existing notions of coordination have typically not been at the level of granularity required to fully explicate *how work is getting done*. More specifically, refer also to Table 7 from Part III of this dissertation, in which individual interviewee statements are linked to dependencies identified in our theory of coordination for commons-based peer production communities. In many of those instances, what the interviewee is describing is essentially the *development* required to effectively contribute to and collaborate with the larger group. Within that development is the sense of *awareness* required to learn the proper skills. For example, while most experienced Wikipedians may find it trivial to defer to a relevant WikiProject to seek out pertinent advice, a newcomer might not know where that WikiProject is, what its purpose is, or even the proper location to ask questions of project members<sup>50</sup>. Beyond these potential structural limitations, there also exist the *cultural* limitations.

For instance, recall from Part III research led by Rughinis (Rughinis & Matei, 2013) in which the researchers examine interactions within a single section of the talk page for the Wikipedia article on Steve Jobs. Ultimately, they find that utilizing *delegated voice* – essentially when an editor adopts the tone of the wise Wikipedian – both increases the perceived legitimacy of an editor’s point as well as the likelihood that it would be attended to. These examples illustrate how the community has adopted social practices that enhance coordination by taking the three pillars of McGrath’s typology of group modes and functions into account, albeit potentially not purposefully. To ensure that these social practices are maintained, which remain invaluable to the onboarding experience of new community members, design for these communities must extend from a deeper understanding of the interactions that comprise the development and encouragement of group members, beyond simply the content they contribute.

Finally, recall from Part I two studies by Morgan et al. (2013) and Viégas et al. (2007) that each attempt to classify the types of coordination work that occurs on WikiProject and article talk pages, respectively. While Viégas found

---

<sup>50</sup> Regarding the proper location to ask questions, it is true that most WikiProjects will respond to questions asked on their Wikipedia talk page. However, some new projects, the Teahouse for instance, utilize a special question gadget to record questions on a different page (see <https://en.wikipedia.org/wiki/Wikipedia:Teahouse/Questions> for an example). In this case, the Wikipedia talk page is typically used by project members (i.e., the Teahouse hosts) to discuss issues related to the Teahouse itself, where the *separate* questions page (linked above) is intended to be used by newcomers seeking helpful advice on editing Wikipedia. The structural shift is not a dramatic one, to be sure, and as the Teahouse is uniquely aware of the need to be clear to newcomers the Wikipedia talk page for the project has a distinct banner stating the purpose of that page, but to newcomers it may still pose a cognitive hurdle in attempting to find solutions on that *or other* projects.

This example is mainly given as a concrete indication that processes may differ between projects, and as the techniques and tools available to project maintainers and the Wikipedians who rely on them increasingly fracture, as is likely to happen as each project matures with the techniques and tools available to them, it may become increasingly difficult for newcomers to navigate those fractured processes effectively. In this instance, both development and cultural awareness would become even more critical to recognize as dependencies that should be considered in future design – one of the primary tenets of this dissertation.

that the predominant work-driven requests on article talk pages were requests for collaboration (i.e., “Come help me fix this article please”), the predominant request on project talk pages were requests for opinion (i.e., “I want to update the template image, what do you think?”). Each of these findings suggests that while the type of discourse may be static between global talk pages, the context and needs of communication differs dramatically between them. In other words, any tool-mediated solution that lacks sensitivity to these distinctions would have the potential to either solve the right problem in the wrong place, or more problematic, to introduce barriers of communication given incorrect assumptions about the context in which the solution is being introduced.

**Table 10. Overview of prior coordination studies.**

<b>Study</b>	<b>Subject</b>
Beschastnikh et al., 2008	Quantitative analysis of all policy citations on Wikipedia talk pages
Crowston et al., 2007	Inductive case study of 1690 emails from 308 contributors to three FLOSS projects
Forte et al., 2012	Mixed methods analysis of interviews with 20 Wikipedians
Krieger et al., 2009	Qualitative analysis of 16 Wikipedians
Kriplean et al., 2007	Mixed methods analysis of policy citations on all Wikipedia talk pages
Morgan et al., 2013	Mixed methods analysis of 788 work-related WikiProject threads from 138 projects
Rughinis & Matei, 2013	Qualitative analysis of individual talk page section
Viégas et al., 2007	Manual classification of talk pages for 25 Wikipedia articles

#### **4.2.2 Reviewing Prior Works through the Lens of a Theory of Coordination for Commons-Based Peer Production Communities**

The goal of the previous section was to revisit prior research with a focus on design for peer production communities, in part to provide a historical perspective for the foundations of the design work that follows. This section aims to continue that review, but to instead situate the assumptions and findings of those prior works using our theory of coordination for commons-based peer production communities, recalling my earlier statement from Part I of this dissertation citing the need for a precise instrument to *describe* ongoing interactions within peer production communities as well as a means of having the results of that analysis inform potential design interventions that could improve the coordinative actions being described.

For example, consider research by E. Gilbert (2012) introduced above in which duplicate responses were identified and avoided in Twitter social networks by resolving the “triadic awareness” problem. From the perspective of the originator of the message (i.e., the individual using the tool), such a tool provides an external means of ensuring that the *product* of the interaction – namely, the tweet – adheres to the social norms valued by the community (i.e., try to avoid spamming your social network with information they’ve already received. From the perspective of our theory of coordination for commons-based peer production communities introduced in Part III, this small act involves a deceptively complex network of dependencies traversed to ensure that the interaction is properly mediated. First, the

tool must be aware of the social networks of not just the sender of the original message but also the social networks of every user to which the originating sender is also linked to, indicating *group awareness* within the network (i.e., “monitoring at the *structural* level of generic group connection). Once the network is known, the tool must then determine whether any user in the originating senders’ network has received the same content from any other user in the destination users’ networks. For instance, if User B follows User A and User C, but User A does not follow User C, it remains that User A will not know what User C has heard from User B and, consequently, it is possible that User B will receive duplicate content from Users A and C. In this case, the tool must be aware of social networks two degrees away from the originating sender (previously established as *group awareness*) as well as the content of prior messages that were sent, indicating *social awareness* (i.e., “lurking project communication,” or “knowledge to ask the right question of the right group member”). Given the knowledge of who has previously seen what content, thereby resolving the social translucence problem which gave rise to the potential for the problem in the first place, the originating sender has the option to refine, revise, or refuse to send the message to those whom may have already received it, in this case indicating a *production activity* dependency (i.e., “content creation,” or “completing work to further group goals” – in this instance, the goal being the successful distribution of unique content respecting the social norms of downstream members of one’s social network)<sup>51</sup>.

Another example is provided by a study led by Cosley (Cosley et al., 2007) that introduced the design and evaluation of SuggestBot, a tool within Wikipedia that gave recommendations of articles to edit based on prior interactions with Wikipedia. For instance, if User A spent a large portion of time editing articles related to Human Computer Interaction, SuggestBot may recommend editing articles related to HCI along a variety of lines, from adding sources, handling orphan articles, or expanding stubs (Figure 14 shows an example of a list of articles recommended for editing). Similar to the Twitter triadic awareness tool described above (E. Gilbert, 2012), the role SuggestBot plays in editing activities, from the perspective of our theory of coordination for commons-based peer production communities, is more involved than may be apparent from a cursory glance. Specifically, from the perspective of the user who is requesting article recommendations, the initial request involves adding a specific template to a page in the User namespace, indicating both a *social awareness* dependency (i.e., “knowledge required to ask the right question of the right group member”) as well as a *social production* dependency (i.e., “communication required to coordinate activities around non-group artifacts”). Once the recommendations have been received, the user then has the opportunity to self-select any of the tasks that were recommended of him or her and to potentially complete that task, indicating a *role identification* dependency (i.e., “task identification, task selection,” or “bringing *existing knowledge* to select coordination/production activities”). Finally, once the user identifies a desirable task, it is possible that he or she needs to enlist others’ help to complete that task (i.e., *social production*, or communication required to coordinate production activities), potentially followed by committing the change requested by the SuggestBot tool (i.e., *production activity*, activity to further group goals), potentially concluding by posting an update to the SuggestBot provided task list that the activity suggested was completed (i.e., *coordination activity*, completing work required to coordinate group activity) and requesting that another

---

<sup>51</sup> Another example of this type of interpretation of interactions can be found in one of the scenarios utilized to evaluate our theory of coordination for commons-based peer production communities, included as Appendix E.

Wikipedian review that work to ensure the quality was high enough (i.e., *social production*, communication regarding production of an external artifact).

Both the examples given above are intended to be illustrative but not exhaustive. It should be clear that our theory is not intended to provide a means of fully recreating the myriad interactions possible within commons-based peer production communities, but that it aims to provide a common language that members in, designers for, and researchers of those communities can utilize to better identify salient moments of information transfer or coordination activities. Once identified, interpreting those moments through the lens of our theory of coordination for commons-based peer production communities affords those members, designers, or researchers the ability to expand and explore the interaction; to more clearly identify where those dependencies *function*, where they *falter*, and where they *break down*. Once identified, these moments then more concretely suggest where design interventions may have the most positive impact, supported by the shared vocabulary of our theory of coordination.

Ultimately, this review aims to more effectively situate prior works with regard to our theory of coordination for commons-based peer production communities. As illustrated in Part I of this dissertation, there is currently little agreement on what constitutes coordination in peer production communities with regard to the *actual activities being performed*. As a result, it has been difficult to evaluate one approach over another aside from purely quantitative measures which are not cross-community, qualitative measures which are not cross-vocabulary, or post-hoc theory-based explanations which do little to provide detailed actionable directions for future design interventions. By being sensitive to these concepts, and by addressing an expansive rather than constraining approach to coordination, I aim to design an application to enhance group coordination. In the following sections, I introduce both the architecture and prototype of such a collaborative application.

### 4.3 SYSTEM ARCHITECTURE


With the Virtual Team Explorer (VTE) and its underlying architecture, I have created a collaborative platform that facilitates interactions at multiple levels without constraining the social and behavioral aspects of existing practices in Wikipedia. Accordingly, in our development process, I focused on four primary requirements. First, the means in which data for the application is stored would need to be accessible, modifiable, and familiar to users both from within Wikipedia as well as those accessing that same information through the application. Second, the platform for deployment would need to be familiar and stable, as well as easily extensible given the varied needs of the broad spectrum of users and use cases. Third, multiple streams of data from disparate information sources, including wiki pages, the Wikipedia API, and external APIs, would need to be supported. And fourth, the functionality of the tool would need to extend beyond the current functionality of the MediaWiki platform to both increase the novelty of the application and potentially its adoption, as well as to enable new forms of communication through and interaction with that platform. The need for these requirements is situated not in the need to resolve any immediate community members' dependencies as may be defined by our theory of coordination for commons-based peer production communities, but to ensure that the introduction of a new tool does not upset the existing balance by introducing new, *unresolved* dependencies, either by fracturing information streams in an untenable way, by removing current

processes outside the grasp of those who would chose not to use a tool-mediated approach to coordination, or by enacting a procedural change that contradicts the underlying culture of the community. The following sections explore in more detail each of these design requirements and the manner in which each requirement is addressed in the VTE.

**Open tasks**

Please feel free to **add to this list!**

To **edit** this list, **select here** → [Tasks](#)

 Here are some **tasks awaiting attention**:

- **Assess** : Please tag and **assess** breakfast-related articles with the project's banner:  
{{WikiProject Breakfast|class=|importance=}}
- **Expand** : Champagne breakfast, Eggs Benedict, Elevenses, History of breakfast, Lox, [Maypo](#), Mini Cinnamon Churros, Pancake breakfast, Suhoor
- **NPOV** : [Breakfast](#)
- **Photo** : [Maypo](#), [Milk toast](#), [Creamed eggs on toast](#)

Figure 15. Example Task List from WikiProject Breakfast.

### 4.3.1 Data Structures for Cooperative Design

One chief concern in the development of the VTE was to ensure parity between data representation within the application and the presentation of that same information within Wikipedia, while still facilitating a depth of experience that might not be feasible with the MediaWiki platform. For example, many projects have adopted the practice of including a task or TODO list on their project pages, with each task typically including a title and description, with others including additional information such as links to specific articles, task owners, task priority, or many other potential metadata (See Figure 15 for an example task list from WikiProject Breakfast).

Accordingly, I would need to adopt a data structure that fulfilled the following criteria:

- 1) It would need to be in-wiki so that full adoption of the tool would not be required for any user that tried to experiment with it (i.e., the task list would be human-editable by project members not utilizing the VTE, and tool-editable by members who were);
- 2) It would need to support representations of arbitrary metadata either in-wiki or within the VTE (i.e., as some project lists tasks with just a title and a description but others also include links and priority, the data structure would need to facilitate each of these types of representation);
- 3) The data structure would need to be able to support data storage in-wiki that would not be represented unless that object was displayed within the VTE (i.e., tasks may be represented hierarchically which may be non-trivial to accomplish in-wiki, but if the data structure supports sub-tasks which can be ignored in-wiki, the VTE can easily represent them in tree form); and

- 4) the data structure would need to be resilient to malformed data or human error so that tool-representations of the data are still feasible regardless of data quality (i.e., if an editor mistakenly enters malformed data, the VTE should be able to identify and attempt to fix that data, and the in-wiki representation should alert the user of the potential problem).

To address these primary issues, I chose to adopt the MediaWiki Module<sup>52</sup> to store and structure data. This choice came with a number of benefits. First, as it is a MediaWiki supported technology, even project members with no prior knowledge or experience would be able to modify or contribute to data content without issue. Second, the ability to parse the structure in Lua provides us with the ability to validate and format data, ensuring that malformed input can be appropriately managed or ignored. And third, as the module invocation provides a structured way of storing unstructured information, it provides the ability to store arbitrary data (i.e., task lists, member lists, lists of lists, etc) utilizing the same module. In other words, it allows project members to store arbitrary lists of attribute-value pairs, validate those pairs, and represent within Wikipedia any of those arbitrary values in any manner they desire – for example, a simple modification to the module code would allow things like priority to be represented as color-coded urgency if it was an integer, or as the raw string if it was any other data type.<sup>53</sup>

### 4.3.2 Platforms for Deployment

Beyond data structures, the platform for deployment was another concern during the design of the VTE. Prior tools have utilized proxied versions of Wikipedia to overlay their own user interface on top of Wikipedia pages (for instance, WikiTasks in Krieger, Stark, & Klemmer, 2009 and Re:Flex in McDonald, Gokhman, & Zachry, 2012); however, this approach has its own shortcomings. Namely, proxied users are forbidden from making edits to Wikipedia to reduce the risk of vandalism from anonymous, untraceable editors. As one of the primary goals of the VTE would be to facilitate project-related tasks that would require updating a variety of Wikipedia pages, this approach was not a feasible solution. A second option was to have a stand-alone tool that would allow users to log in and complete project work without requiring a browser, but the complete disconnect from the WikiProject was seen as potentially not conducive to the project work that I was trying to facilitate. As prior research has shown (Hill, 2013), along the spectrum of novelty versus familiarity, success tends to favor novelty of innovation, familiarity of structure (see Figure 16).

Ultimately, the VTE was created as a User Script within Wikipedia. These User Scripts provide a means for editors to create custom JavaScript that can be inserted into any client browser once that script is activated, allowing for a powerful and simple means of creating our own overlay on existing Wikipedia pages. Similar to above, this selection came with a number of benefits. Once again, alignment with existing MediaWiki technologies and practices was desirable, as I felt that it would increase the potential for the tool to be adopted on even a modest scale. Additionally, along with familiarity of structure is ease of installation. As a user script, the VTE can be

---

<sup>52</sup> See <http://en.wikipedia.org/wiki/Wikipedia:Lua> for more information on modules.

<sup>53</sup> More information on this data structure can be found in Appendix H in this dissertation, or view Module:ListMaster and Module:ListItem for more details on usage.

“installed” by adding a single line to a special Wikipedia page<sup>54</sup>. Further, since the VTE exists within Wikipedia<sup>55</sup>, it can be freely copied, modified, or distributed, providing individual WikiProjects, testers, or curious editors the ability to test or expand its functionality in a simple manner. Finally, situating this tool within the familiar confines of the MediaWiki platform was a means of increasing opportunities for appropriation by the larger community – “The ongoing, incremental adaptation of interactive technologies is inherent to the emergence of practice, and practice is inherently shared” (Dourish, 2003).

		Innovativeness of Goal/Product	
		Familiar (P1=Y)	Novel (P1=N)
Innovativeness of Process/Tools	Familiar (P2,3=Y)	TDEP GNE Nupedia	Interpedia  Everything2 h2g2
	Novel (P2,3=N)	Wikipedia	The Info Network

**Figure 16. Diagram from (Hill, 2013). While Everything2 and H2G2 are still active, our platform selection aimed to mirror Wikipedia on the Novelty vs. Familiarity spectrum. Reproduced with permission from the author.**

### 4.3.3 Extensibility of Approach

Many tools have been created to facilitate work in Wikipedia. Tools exist to facilitate the process of identifying open tasks, to automate frequent edits, to visualize editor activity, to alert for potential vandalism, and to enable editors to more effectively contribute to the encyclopedia, among a plethora of others. In the design and development of the VTE another primary concern was the ability of the architecture to support the aggregation of multiple streams of data in an extensible manner to more effectively support emergent uses not originally designed for. This currently includes data from three distinct sources: the MediaWiki API, Wikipedia pages themselves (i.e., the modules that facilitate in-wiki data manipulation as well as the many talk pages which provide a means of communication and discourse, both within and outside the VTE), and a custom API I created that provides an interface to myriad project and editor activities, described further in the following section.

#### Yet Another API

While the MediaWiki API provides a means to retrieve a vast amount of information about activity within Wikipedia, implementation of the VTE required streams of project-specific data that would be tedious and time consuming to retrieve through that approach. For instance, to retrieve a list of edits to articles under the scope of a

<sup>54</sup> See [http://en.wikipedia.org/wiki/Wikipedia:User\\_scripts](http://en.wikipedia.org/wiki/Wikipedia:User_scripts) for more information on installation instructions and to view a list of currently available scripts.

<sup>55</sup> Code and documentation for the VTE can be found at [https://en.wikipedia.org/wiki/User:Md\\_gilbert/vte](https://en.wikipedia.org/wiki/User:Md_gilbert/vte).

given WikiProject within the last month would require a tool to both identify that list of articles as well as fetching the recent revision history of every single one of them. For a WikiProject of any significant size, this simple task would take a significant amount of time. As our aim for the VTE was to create an interactive tool that provided a consistent user experience across all WikiProjects, I required a means of retrieving that and similar data in a manner that would allow data to be retrieved in seconds rather than minutes. To accomplish this, I utilized the Wikimedia Tool Labs<sup>56</sup> to create a cached version of the metadata for all users, pages, and revisions within the English Wikipedia. This approach facilitates the ability to deliver aggregate level data about interactions with Wikipedia that can be grouped by user, page, or week, delivered more quickly in a more customizable form than either the MediaWiki API or the Tool Labs servers would support. Beyond the basic cache of revision metadata, I additionally collect and provide access to the following:

***WikiProjects*** – Individual WikiProjects are identified through the union of pages matching a particular string (i.e., all pages within the Wikipedia namespace starting with “WikiProject\_”) as well as through a category search (i.e., pages in the Active\_WikiProjects Category). This allows us to identify more traditionally named projects, such as WikiProject Breakfast, as well as those that act as a project but do not follow the naming scheme, such as the Department of Fun.

***Pages under the scope of each WikiProject*** – These are identified through a recursive traversal of the category structure of individual projects. This allows us to collect all pages under the scope of, for instance, WikiProject World’s Oldest People, which will additionally all be under the scope of WikiProject Biography as the former is also within the category structure of the latter.

***WikiProject members*** – Project members are operationalized as users who have added a link back to their user page on any WikiProject page, sub-page, or any page transcluded on those pages. This allows us to include member or participant lists for projects that utilize a module or template for that purpose. This, combined with the aggregate cache of all user activity across the entire English Wikipedia, allows tool designers to utilize and differentiate between both an explicit as well as procedural definition of group membership.

***User reverts*** – Reverts are collected through identifying matching checksums within any page across the entire history of that page. Multiple reverts can be consecutive, nested, or be interchanging, allowing a simple means of picking out edit wars from any page within Wikipedia.

Each of these data points, among many others<sup>57</sup>, provide a simple and open means for developers and designers to utilize this information in an emergent fashion to more effectively support appropriation by the larger community.

---

<sup>56</sup> See [https://wikitech.wikimedia.org/wiki/Help:Tool\\_Labs](https://wikitech.wikimedia.org/wiki/Help:Tool_Labs) for more information.

<sup>57</sup> For more details on what the API provides, as well as the code to run it, please see <https://github.com/mdgilbert/node-reflex>. For details on all data that was collected and the code to do so, including details on creating an identical local cache, please see <https://github.com/mdgilbert/wiki-tools>.

### **4.3.4 Synchronous-Pedia**

A final primary concern was that the architecture facilitate synchronous communication between clients. This desire was mediated by the joint needs to not only promote a sense of activity within WikiProjects, but to promote that same sense of activity in the VTE itself. As the above API was implemented as a Node.js application, implementing a synchronous connection utilizing an open socket between the server and each client is a trivial practice.

Utilizing this open connection, it becomes possible for designers and tool developers to push notifications to individual clients, to WikiProject members, or to push notifications globally. Far more than allowing for the implementation of a simple chat server, this type of connection allows every client to communicate state information to all others, creating a more vibrant community of users and facilitating more organic group-based collaboration than is currently afforded by the MediaWiki platform. The novelty accompanying this increased utility, as well as the simplicity of implementing new features, is intended to highlight the potential that future tools may have on affecting the working practices of current Wikipedians, and to point towards affordances that new tools may additionally be able to adopt in the future.

## **4.4 SYSTEM DESIGN**

The architecture described above was created to enable and support a novel means of tool-mediated coordination within Wikipedia. VTE builds upon the existing MediaWiki platform, but further, facilitates a user experience that provides greater flexibility and opportunity for information transfer and organization than the native platform allows for. In the following sections, I lay out the justification for such an application, the opportunities afforded by that application, and the way in which those opportunities have been implemented.

### **4.4.1 Supporting Collaborative Work**

Much of the prior research into coordination practices within online teams concludes with calls for future action, implications for design interventions that have the potential to improve the means that coordination occurs within these distributed teams. Such an intervention would need to be integrated into the community in a manner that fosters future growth and development along with wider adoption (Dourish, 2003). It would need to be sensitive to the current cultural and social practices of the members of that community to ensure that member development is not negatively affected by its introduction (Kriplean et al., 2007). This intervention, beyond merely influencing or improving means of producing new content, would need to support the well-being of the group itself, as well as to foster the development of the members of that group (Forte et al., 2012). I present the VTE as a prototype application intended to address the requirements this prior research has highlighted. In doing so, I hope to both improve the ability of WikiProject members to complete project work as well as to further the discussion regarding coordination in online teams. In other words, as previously identified requirements are met, as previous dependencies are resolved, what implications are there for future theory development around coordination within

distributed teams, and how might that theory development inform future iterations of tools designed to serve that function?

## **Project Browser**

The project browser is the entry point to the VTE for first time users (Figure 17). This browser provides a simple means of exploring currently active WikiProjects, enabling users to quickly and easily find the most (or least) active WikiProjects within the last 30 days. The level of activity can be defined as the number of edits to project pages and sub-pages along with the corresponding pages in the Wikipedia Talk namespace (typically a good indicator of an active project, and a metric frequently cited as defining a successful project), the number of pages edited within the scope of the project, the number of edits to those pages, or the ratio of edits per page. The intent behind this browser is to enable Wikipedians, regardless of prior implicit or explicit membership in projects, to easily browse through projects that they might be interested in contributing to. It provides a light-weight interface through which to explore the myriad types of topics and tasks that are currently available to interested editors, and a simple means to drill into what those topics and tasks may entail.

From the perspective of our theory of coordination for commons-based peer production communities, this browser serves multiple functions. First, for the community member using the tool it primarily resolves a group awareness dependency. Specifically, knowledge of group activity at the holistic, structural level is necessary to understanding how active the group is, which has been stated by many of the interview subjects who participated in the research leading to this dissertation as a prime factor that most Wikipedians use to determine the overall health of a project, in practice. And second, should the community member using the tool elect to explore a WikiProject represented in the project browser, it indicates the potential resolution of a previously unresolved group identification dependency. Specifically, as group identification is contraindicated by a lack of group identification (i.e., membership), the community member's decision to explore the nature of that project indicates potential interest to align their efforts with those of the group.

Search for a WikiProject in the box above, or select from the list of most active WikiProjects below to continue.  
 (Projects below represent the most active WikiProjects by edits to member pages within the last month, limited to those with at least 30 edits)

Sort by edits	Sort by pages	Sort by ratio	Sort by project edits
<b>WikiProject Good articles</b> Project Edits 633 Edits 61339 Pages Edited 10291 Edits per page 5.96	<b>WikiProject Biography</b> Project Edits 485 Edits 35295 Pages Edited 15739 Edits per page 2.24	<b>WikiProject Wikify</b> Project Edits 2 Edits 32292 Pages Edited 8075 Edits per page 4	
<b>WikiProject Featured articles</b> Project Edits 0 Edits 23206 Pages Edited 3165 Edits per page 7.33	<b>WikiProject United States</b> Project Edits 382 Edits 16420 Pages Edited 5938 Edits per page 2.77	<b>WikiProject Chemicals</b> Project Edits 10 Edits 6501 Pages Edited 2082 Edits per page 3.12	
<b>WikiProject India</b> Project Edits 77 Edits 6451 Pages Edited 3165 Edits per page 2.04	<b>WikiProject Wikipedia-Books</b> Project Edits 0 Edits 5811 Pages Edited 2212 Edits per page 2.63	<b>WikiProject Former countries</b> Project Edits 1 Edits 4922 Pages Edited 986 Edits per page 4.99	

Figure 17. The project browser within the VTE provides an entry point for new users to explore active WikiProjects along a number of metrics.

## Member Browser

One of the most identifiable components of any WikiProject, and frequently one of the least structured, is project membership. The very idea of what constitutes a member within a WikiProject is not settled. In Morgan et al. (2013), membership was operationalized as a user link added to a project page or sub-page with titles like *Member* or *Participant*. Other research has measured membership through the use of WikiProject user boxes or categories on user pages, or by simply equating an edit to a WikiProject page or sub-page as an implicit signal of membership<sup>58</sup>. In the VTE, to ensure that our notion of membership was as inclusive as possible, I distinguish between three types of membership. First, users who place links on project pages, sub-pages, or pages transcluded on either of those can be assumed to be explicit members of a project. The granularity of membership data returned by the API described above is at the level of any user link added to any project related page at any time, including links added by third party users, so the choice of how inclusive this definition of explicit is remains at the behest of designers and future tool developers. My interpretation of these explicit project members includes all user links, excluding those placed on WikiProject talk pages. Second, I include but differentiate between those prior explicit members and users who have made edits to any WikiProject page or sub-page, again excluding talk pages, as these types of interactions may signal a more strong willingness to contribute to, if not identify with, that group. Third, and finally, I provide access to users who have made contributions to pages under the scope of a given WikiProject – this last group made accessible for purposes of transparency and potential outreach rather than as an organization mechanism for the users themselves. This would allow, for example, project members to identify prolific non-member contributors within a topic area to more directly engage or interact with those individuals.

This *procedural* notion of membership has two primary strengths. First, it facilitates the appearance of a more active and lively group, as the list is based on recent interactions between group members rather than static wiki-text.

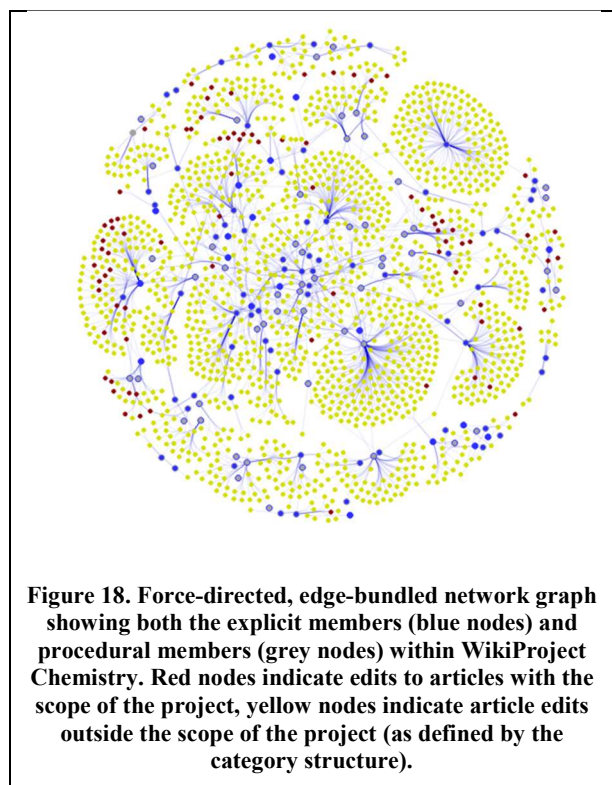
<sup>58</sup> See [https://meta.wikimedia.org/wiki/Research:Visualizing\\_WikiProject\\_Activity](https://meta.wikimedia.org/wiki/Research:Visualizing_WikiProject_Activity) for a concise summary of how membership may be interpreted.

Second, maintenance of that list can be completed in a tool-mediated fashion, ensuring that stale information or long-retired project members are not prominently displayed. And third, it encourages exploration of both the current explicit members of a project as well as the second and third levels of members described above, creating a concrete means of both identifying the users at work within each group as well as explicating the details of the work that they are doing. For example, Figure 18 illustrates the member graph for WikiProject Chemistry. In this graph, blue nodes represent explicit group members, grey nodes represent non-members, red nodes represent articles within the scope of that project, and yellow nodes represent articles outside the scope of the project. Beyond the simple utility of being able to visualize project members and varying levels of activity at a glance, the network approach to visualizing membership additionally provides a simple way to identify potential candidates for membership who may make a strong positive impact on the group (grey nodes who interact with a large number of red nodes) or even articles which are currently not under the scope of the project which potentially should be (red nodes interacting with a large number of yellow nodes).

Similar to the project browser above, the member browser provides a powerful tool to resolve potential group dependencies from the perspective of our theory of coordination for commons-based peer production communities. First, by displaying the network graph of all WikiProject members along with all non-WikiProject members, as well as both the articles within and without the scope of the project that either has edited, the community member who utilizes this tool is provided with a very fine-grained understanding of *who* actually does *what* within the project, indicating *social awareness* (i.e., knowledge of others' roles, knowledge to navigate social structures). With this dependency resolved, with this social awareness in place, the community member using the tool has the capacity to direct questions to the Wikipedian who appears most capable of answering that question in the most appropriate manner. For instance, if the user looking for someone to help fill out infoboxes for another article, it may be in their best interest to first, find someone who looks like they have worked with others extensively (easily identified in the network graph by the position of the blue and grey nodes) and second, they may want to seek out someone who has also edited user and user talk pages recently (or potentially less so, WikiProject talk page), showing a more explicit level of communication with other community members. By satisfying those criteria, community members using the VTE have the capacity to direct their attentions in a way that is potentially more fruitful than scanning the participants lists within WikiProjects or making judgments based purely off one individual or one page's change logs – the information presented in aggregate provides a much more complete picture while still allowing users to drill down into individual pages and editors.

Second, the member browser may also ameliorate the *group awareness* dependency more generally. As group awareness entails basic knowledge of both the members of a project as well as the content under the scope of the project, the ability to immediately identify both the most active users, the most active articles under development provides a quick way to build a mental model of the structure of those users, articles, and the interactions between them (more specifically, group awareness entails “developing knowledge to complete coordination / production activities,” which this type of awareness would engender. Further, the ability to quickly determine both whether group members were primarily editing articles not under the scope of the project or whether non-group members were editing articles under the scope of the project could give rise to further downstream dependencies. For instance,

in the case of the former it is possible that the article *should* be categorized under the scope of the project in which case the knowledge of group structure (group awareness) could give rise to the activity of categorizing the article (production activity). In the case of the latter it is possible that knowledge of group structure (again, group awareness) could give rise to sending an invite to officially join the WikiProject (in this case, group identification, indicated by the group member's outreach or recruitment activity to other community members). Finally, given the information required to select an article to edit, a user to communicate with, or a job to complete, in the enactment of any of the above the *role identification* dependency is implied in that the user electing to complete that task has aligned with the role each task may entail (e.g., content producer, user outreach, mentor, etc).



## Task Browser

A task browser would provide project members with a structured, trackable means to record desired tasks, assign them to project members, and to track those tasks from creation to completion. Such a system would clarify individual efforts, informing editor reputation and potentially increasing group bonds among project members.

Our implementation of a task browser in the VTE utilizes the module system described above, further expanded in Appendix H of this dissertation. Using this approach I am able to create task lists, or any other list, by invoking individual ListItem modules that can have any arbitrary data assigned to them. This allows us to, for instance, define a task using wiki markup which includes a created date, a priority, a title, and an owner, as is visible at the top of Figure 19.

Additionally, this same ListItem invocation could include subtasks, estimated time remaining on that task, or a due date for any individual item, relying on the module invocation to safely ignore extraneous data from the in-wiki representation of the table, but still providing that data to external tools such as the VTE. Take, for example, a sample of the text used to instantiate the task list shown in Figure 19:

```
{#{#invoke:ListMaster|printTable|
  style=table|
  display=title,description,priority|
  ...
  {#{#title=Vandal watch for [[Mobile...]]|
    description=This article has been vandalized a
      few times in the last week, need help
        monitoring|
    priority=0|
    created=17:54, 1 April 2015|
    subtask0=Comment on the Talk page to let me
      know you're on the task
  }}
}}
```

This additional information can then be integrated into a single user experience. The VTE task list in Figure 19, for instance, includes a due date which has already passed for the first item, indicated by the red background color for the first column. The fourth column, listing the number of comments on that task, represents the number of comments on a talk page section created explicitly for that task by the VTE, ensuring parity between tool-mediated interactions with the task list (via the VTE) and traditional human manipulation of those lists within Wikipedia. Further, this additional information can be simply and uniformly represented to the user when viewing task details, allowing a more structured approach to visualizing and interacting with arbitrarily complex data.

This integration of multiple streams of information into a seamless user experience highlights one of the primary goals of the VTE – to create an opportunity for enhanced collaboration of group members without diminishing the contributions of non tool-mediated editors (i.e., those not using the VTE) or removing opportunities for enculturation or socialization of group members (Kriplean et al., 2007). The goal for this application, then, is to reduce the operational cost of retrieving disparate information from these multiple streams of information without diminishing the social experience that the underlying platform affords.

Once again, from the perspective of our theory of coordination for commons-based peer production communities, the task browser affords a simple means of resolving a few important dependencies for peer production work. First, the act of *recording* any task completed under the scope of the group constitutes *coordination activity*, in that it entails activity to sustain the group through maintaining a shared view of both the work that needs to be completed, and the work that has been finished (or updated, or blocked – the actual work completed is less crucial than the fact that the current state of the work under the project is accurate and shared). Additionally, this first dependency resolved implies a preceding dependency, that of the *coordination* or *production activity* that was updated in the task list (i.e., the task that was to be completed – for instance, “Update the article on Cats” would be a reasonable task

item for WikiProject Cats). Second, from the perspective of all other group members, this *coordination activity* dependency would engender a resolved *group awareness* dependency, in that the originating action (updating the task list) would provide the shared knowledge the remainder of the group could utilize to more effectively direct their future efforts. And third, from the perspective of the group members who view the task browser without creating the task, it is possible that the *role identification* dependency may be more directly addressed as the browser is providing explicit information about what needs to be done in the project. Should that group member (or VTE browser) elect to complete that task, they have in effect aligned with that role, albeit temporarily, to complete the task in question.

+ Add task		View: All ▾	Sort: Created ▾		
Apr 7, 2015	4	Style check	4 comments	Md gilbert	
Apr 4, 2015	5	Request for mentor	0 comments		
Apr 1, 2015	0	Vandal watch for <a href="#">Mobile interaction</a>	0 comments		

Title	Description	Priority
Style check	Could someone make sure that <a href="#">Web design</a> conforms to desired style?	4
Request for mentor	I'd like someone to let me know what types of articles may be best for a newbie to work on	5
Vandal watch for <a href="#">Mobile interaction</a>	This article has been vandalized a few times in the last week, need help monitoring	0

**Figure 19. Task browser in the VTE (top) and as displayed in the WikiProject page (bottom), both using the same WikiText to render. The ability to embed arbitrary metadata in task invocations means that a far greater amount of information (subtasks, for example) can be included without complicating the in-wiki representation.**

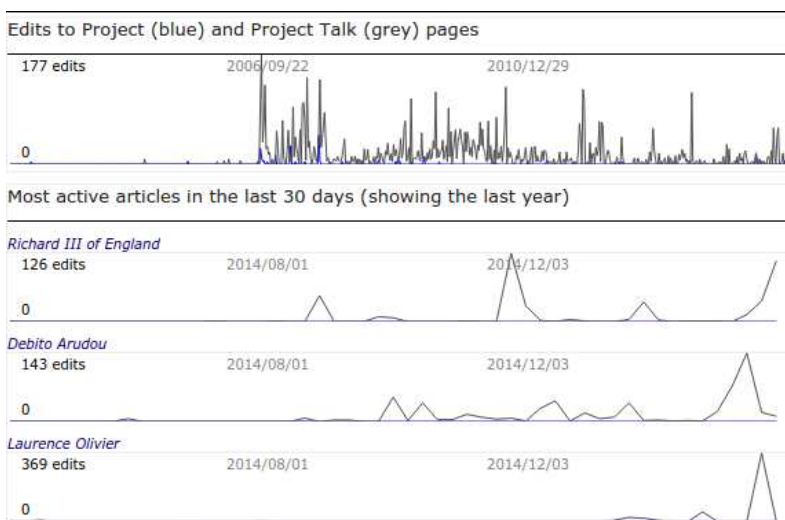
## Article Browser

Finally, the article browser provides a central means for group members to easily browse through the most active articles under the scope of a WikiProject (see Figure 20 for an example from WikiProject Biography). At a glance, this browser is intended to give a high level overview of editing activity of articles under the scope of a given WikiProject for the last 30 days, indicating areas receiving greater attention as measured by edit counts. Such a high level view is useful to pull out global trends across projects as well as to identify current trending topics. For example, during the Ebola virus epidemic during the summer of 2014<sup>59</sup>, one of the most active WikiProjects measured by the ratio of (edits to pages under the scope of a project / the number of pages edited), WikiProject Ghana was one of the top ranked WikiProjects. This was due to the number of edits the Ebola article was receiving, categorized as under WikiProject Ghana through its connection to WikiProject Africa, and was immediately apparent through a cursory examination of project activity.

<sup>59</sup> [https://en.wikipedia.org/wiki/Ebola\\_virus\\_epidemic\\_in\\_West\\_Africa](https://en.wikipedia.org/wiki/Ebola_virus_epidemic_in_West_Africa)

Beyond simple lists of project and article activity, however, the goal for the article browser is to create an open and shared means of creating watchlists for project members. Through these centralized lists, specific articles can be monitored to ensure that relevant activity is easily accessible and useably shareable among project members. Further, utilizing the module structure enumerated in the sections above, these lists can include more varied information than is currently possible using existing MediaWiki watchlists. This may include attaching metadata such as task identifiers to individual events, assigning article owners, creating priorities for activity on articles of varying importance, creating sub-categories of watched articles based on the context of content or activity, or implementing many other solutions as the desires of project members dictate. And as described in the task browser section above, each of these lists would be accessible in both tool-mediated fashion, utilizing the VTE or other applications, as well as within Wikipedia. I leave these extensions to the current implementation as future work requiring further tool mediation, discussed more fully in the sections below.

Once again, from the perspective of our theory of coordination for commons-based peer production communities, the article browser would primarily facilitate resolving the *group awareness* dependency. Similar to the examples above, as the amount of accurate and salient information about group activity is made shared and available, the capacity for group members to act productively is increased. Acting productively, in turn, indicates future downstream dependencies that may be resolved from these representations. As well, if the user of the article browser elects to act based on the information provided, a *role identification* dependency is resolved in that the user in question self-selected to complete a task based on that new representation of information.



**Figure 20. Example presentation of the article browser within the VTE, with the above showing the project and project talk page activity for WikiProject Biography, along with the top three most active articles by revision count for the last 30 days under the scope of the project.**

#### **4.4.2 Methodology: Situating the Virtual Team Explorer within a Theory of Coordination**

While a great deal of research has been completed on coordination within Wikipedia, and a great many tools have been created to facilitate that coordination, there is still little agreement on how this coordination can be most effectively researched and tools most effectively designed for use within peer production communities. I argue that our theory of coordination for commons-based peer production communities is such a tool, providing the means to describe interactions at the levels at which they occur. Specifically, in communities where production is generally the outcome of creativity without recompense, any theory that is used to describe how coordination occurs within that community and must provide the sensitivity and precision to encompass the breadth of the contributors encountered. Further, the goal in providing example dependencies for the systems above was to highlight how, overall, coordination does not exist in isolation from the processes that are being coordinated.

While prior research has conflated meta-work with coordination, or communication within specific channels (i.e., posting to the Wikipedia talk namespace), or posting inquiries of a certain type within specific channels (i.e., posting requests for assistance in the Wikipedia talk namespace of a WikiProject), each of these approaches comprises only a small picture of how coordination occurs (for more examples of interpretations of coordination in prior research, refer to Part I of this dissertation). Through the course of the interviews described in Part III, our research team heard numerous examples of how coordination was enacted within this peer production community. Largely, we learned, coordination was *not* an intentional act within these communities. Rather than through explicit coordination, successful production was more frequently the result of opportunistic contributions leading to incremental knowledge, enculturation through immersion, and group identification, when it did occur, through socialization.

It is through these simple signals, I would argue, that coordination is more effectively enacted within peer production communities. Accordingly, our theory of coordination for commons-based peer production communities aims to increase the capacity for both researchers and designers to both observe and intervene in ways more nuanced than prior approaches have afforded. Through the perspective of this theory of coordination, it is conceivable that a far greater spectrum of interactions may be more effectively typified and interpreted, thereby leading to the creation of more effective tools, whose evaluation will thereby allow this theory of coordination to be further refined for this and other peer production communities.

#### **4.5 EVALUATING THE VIRTUAL TEAM EXPLORER**

After the completion of the interviews with Wikipedians described in Part III of this dissertation, all prior interview participants were invited to complete a survey to evaluate a proposal for the Virtual Team Explorer, as described above. The survey took the form of individual scenarios in which two Wikipedians sought to complete different tasks common to those in peer production communities. In the first scenario, a newcomer to Wikipedia was simply seeking to find a task where he could most helpfully contribute content. In the second, an experienced Wikipedian

was seeking new ways to explore and inspire collaborations among her fellow WikiProject members<sup>60</sup>. The response rate for completing the survey was 100%. A summary of their responses is included below, followed by a brief discussion of those responses.

**Table 11. Results of the survey evaluation of the Virtual Team Explorer prototype.**

	<b>How helpful would the VTE be for the user in the scenario: 1 (least) to 4 (most)? Mean (median, sd)</b>	<b>How likely would you be to use the VTE given the example in this scenario: 1 (least) to 4 (most)? Mean (median, sd)</b>
Scenario 1 – In which a newcomer uses the VTE to find a project to join and articles to work on.	2.556 (3, 0.882)	2.333 (2, 1.00)
Scenario 2a – In which an experienced project member uses the VTE to find and engage community members to join a project.	2.889 (3, 0.928)	2.778 (3, 1.201)
	<b>How useful would the VTE be for the average Wikipedian to track and control common tasks: 1 (least) to 4 (most)? Mean (median, sd)</b>	
Scenario 2b – In which an experienced project member uses the VTE to track and update project member’s status on open tasks.	3.111 (3, 0.782)	

Two points become apparent from the quantitative data gathered: first, the lowest scoring question for the survey was whether the interviewees would be likely to use the tool given the scenario of a newcomer seeking a project to align with and initial tasks to work on. This should perhaps not be surprising as interview subjects were selected specifically for their experience with WikiProjects and, given their tenure within their chosen projects; they likely have less need for elements of the VTE aimed at easing the onboarding experience. And second, given the final question of *how useful they think the VTE would be for the average Wikipedian*, the mean response of 3.111 indicates a clear opinion that our interview participants felt that the VTE could be valuable for future work within Wikipedia. Beyond the quantitative data collected regarding the VTE, interviewees additionally were asked a number of questions to clarify those scores. The table following illustrates some of those questions and highlights key quotes given by survey respondents.

**Table 12. Comments given by survey participants for the Virtual Team Explorer prototype scenarios.**

<sup>60</sup> The survey, including the entirety of both scenarios described above, is included as Appendix G in this dissertation.

Question: The scenario above involves a relatively new editor to Wikipedia, searching for an active group of Wikipedians to more effectively identify both clear and manageable tasks that he could complete, as well as a larger group of editors that he could more actively conspire with. What kinds of additions or modifications would you make to a tool such as the VTE that might make achieving these goals easier for Cameron? What kinds of changes or modifications might make achieving these goals easier for yourself?

Interview Id	Comment
02	<p><i>The tool and the look is awesome. I gave less than perfect ratings because I believe that the tool described in section 1-1 is hampered by not having necessary data feeds to make it actually work<sup>61</sup> ...</i></p> <p><i>In section 1-2 the tool seems like a miracle if you can actually make that kind of report. That is very useful. I have never seen anything like that and would want to have access to a graph which could show me which articles in a WikiProject are highly active.</i></p>
04	<p><i>I think the VTE tool itself would work with regard to allowing an editor to find an active project. From there, assuming the new editor knows about talk pages, they might visit the project talk page to look at recent discussions or they might consider what the project itself has included in its "to-do list". An analytical approach might work in finding a relevant project, but I'm not sure that analytical approach necessary carries through to a large number of people using a tool like that to help find things to do for themselves...</i></p> <p><i>It would be useful if project members could somehow tag articles for VTE attention - "these are the articles with the most activities, but there are these others that project members have tagged as needing attention".</i></p>
05	<p><i>I wouldn't use the tool. I have way too much experience, so I know how to find what I want without fancy tools. That being said, I think it'd be useful for new users who don't have this knowledge and grew up with the Internet always being around.</i></p>
07	<p><i>My concern about Cameron in this scenario is that active articles usually flag controversy, and shuffling new editors into controversial areas can set them up for immediate failure...</i></p> <p><i>I would modify the tool to ensure that Cameron is advised that high activity on articles may sometimes make working on them more challenging and encouraging him to read the talk page before diving in.</i></p>
09	<p><i>I'd like a tool like this that might point me towards more active projects on the project list... but it would still require that someone inside of the team be an active force and self-generate editorial structure. There would have to be some type of central body issuing commands and asking for fulfillment...</i></p>

Question: The scenarios described above involve a fairly experienced editor utilizing a new tool to both structure WikiProject-related tasks as well as to make completing some of those tasks easier. What kinds of additions or modifications to this tool might you suggest that might make Charlie's goals of organizing WikiProject Women writers easier? What kinds of changes or modifications might make achieving these goals easier for yourself?

---

<sup>61</sup> The data feeds described in section 1-1 of the survey do exist, as described in the sections above, although that was not made clear in the description of the scenario.

02	<p><i>About [categorizing articles using the VTE] - There are two activities which Wikipedians do. Content contributors add content, often in a domain, and functionaries provide administrative support. Some of the most popular WikiProjects are for functionaries, and functionaries often have no loyalty to any particular WikiProject. These people may implement template updates, do copyediting, do mediation, close discussions, and the rest. The tool system here would not apply to that large class of users. For people developing content this seems very useful.</i></p> <p><i>About [reaching out to potential project recruits] - Yes this would be a great way to get invitations to people who would be interested in a WikiProject... If this tool were made and had this function, then I think one of the outcomes would be an interest in more WikiProjects making custom invitations to meet their community's particular needs.</i></p>
04	<p><i>The scenario provided certainly had merit but it seems to operate on the premise that the editors in question are only ever members of the one WikiProject and that articles are only categorised into one WikiProject each. In reality, the two people she wants to invite may well be part of a different WikiProject together, like WikiProject Literacy or WikiProject Books. Those higher-level project (larger, broader in scope) mean that people might be collaborating together on a series of articles that might be group under more than one project...</i></p>
09	<p><i>Oh, I like this possible tool. The ability to see a visual representation of user contributions and to explore for new recruits is kind of awesome. I currently have to try doing this by reviewing individual pages for folks who have edited them recently, and to invite or discuss with them on user pages to become more focused or to spread out to other articles that need attention, with limited success. This would at least give me a quick top-down look on who might be recruit-able.</i></p>

Question: In closing, each of the scenarios described above are intended to illustrate how a unique tool can integrate various channels of information to facilitate how WikiProject members can organize and complete their work. Extending from this goal, do you have any comments or criticisms regarding the general goal, or ideas that you feel may be more effective in supporting WikiProject members?

02	<p><i>In my opinion, all WikiProjects should be supported with standardized tools by default, rather than created with no tools in place as is the current practice. If tools are in place, then the alert system that you describe here would be very useful...</i></p> <p><i>I like these tools a lot. If it is feasible to make them then I think they could become part of the standard expected experience on Wikipedia.</i></p>
03	<p><i>I believe this tool could be helpful for WikiProject participants, as described here. I think that the tool will be more effective for experienced Wikipedians than newcomers.</i></p>
04	<p><i>I'd draw attention to comments about ease of sign-up. Some optional tools are very powerful and without knowing exactly what they do, adding user script (like a console edit) might turn some people off. A easier opt-in/opt-out check box like other tools in the "editing tools" list would be better...</i></p> <p><i>My point, broadly, is that handing very powerful tools (even collaborative ones) to people with no "qualifiers" (anyone can use them and "anyone can edit" with them) will mean that the package comes in for significant scrutiny. But, broadly, the benefits far outweigh the risks in this instance.</i></p>
06	<p><i>Overall, I think your efforts to improve the participation in wikiprojects are fantastic. Thank you for doing this work.</i></p> <p><i>I continue to suspect that much of the vandalism that we see represents potential wikipedia editors</i></p>

	<i>who are interested in wikipedia enough to figure out how to make an edit and save it, but are misguided in their efforts. While I try not to bite the noobs, I wish there was a way to quickly and rapidly subvert their vandalistic tendencies toward a productive end...</i>
08	<i>The general goal of facilitating cooperation seems like a good one... [A] common task that was not mentioned in this survey is the coordination of media creation like photos. Particularly for geographically-centric WikiProjects like WikiProject California, WikiProject Houston, etc, there are usually lists of requested photos - usually on articles that are already geotagged...</i>
09	<i>VTE sounds like a cool toy, but the name needs rebranding, and maybe there need to be more "new user" tutorials that we can pass out to new folks? Quickie guides that gradually help folks to enter in without the sharp learning curve or the occasional sharp reprimand from experienced users frustrated by perceived vandalism? Lowering the bar for new users and providing more experienced users with more toys to play with sounds like a good combination?</i>
10	<i>The tool sounds very promising!</i>

Ultimately, the comments left by survey respondents echo the sentiments expressed in the sections above. First, existing tools for coordination within WikiProjects are fractured, with each project needing to adopt, adapt, and maintain each of the tools they elect to use. Second, that any coordination-minded tool utilized by WikiProjects should be sensitive to both newcomers seeking basic project information as well as more experienced Wikipedians seeking more streamlined means of finding relevant information about the projects they interact with, simple means of recording and updating the status of those projects, and satisfying means of interacting with the projects in general. And third, if the VTE as described was to provide a cohesive experience for both new and experienced Wikipedians, it would likely have a positive impact on participation in WikiProjects as well as on the work those projects attempt to accomplish.

## 4.6 DESIGNING BEYOND MEDIAWIKI

The MediaWiki platform has repeatedly been shown to be an effective tool for building consensus and nurturing an increasingly complicated policy environment (Forte, Larco, & Bruckman, 2009; Viégas, Wattenberg, & McKean, 2007). This environment is both constrained by the code that makes up the platform as well as informs the direction of future technological interventions within it – “In other words, artifacts have politics... and code influences the development of policies and norms,” (Forte, Larco, & Bruckman, 2009). Beyond the development of policy, though – a process which involves reducing a cacophony to something more resembling reasonable thunder – there remains the need to coordinate the efforts of distributed group members, each of whom contribute with different motivations, differing talents, varying schedules, and unique histories informing the content of their contributions. Beyond the how-things-should-be-done arguments of policy aimed at building and nurturing a healthy community with codified and understandable norms, coordination at this level must contend with the vagaries of these differing motivations, talents, and schedules.

I suggest that, at this level, the MediaWiki platform may not be as effective at facilitating coordination compared to the functionality that an ecosystem of free and open external applications can provide. The VTE and the underlying architecture presented above is, I contend, one such application. The structural separation of content among namespaces remains largely symbolic; the structure of the dialogue that occurs on the numerous talk pages among them is driven more by convention than constraint. Through the development and refinement of tools that support emergent processes in these distributed teams, I hope to facilitate an environment in which prior research-generated theories can readily inform future design. And, in turn, future design and the study of its use may more effectively illuminate the nature of complex socio-technical systems, such as Wikipedia.

## 4.7 NEXT STEPS

Coordination within commons-based peer production communities is a subject that has been broadly studied in the last five years. From the research cited in the review above to the recently launched WikiProject X<sup>62</sup> and Wikipedia Co-op<sup>63</sup>, which aim to better understand and design for the project space as well as to provide mentorship within the community at large, respectively, design-oriented researchers have sought to better understand how coordination works within wiki spaces and to, accordingly, improve it.

The VTE, currently presented in prototype, comprises a collective attempt to transform that research into practice. Continued development will be necessary to, first, ensure that the architectures presented here are suitable to the needs of those they are intended to serve and, second, to evaluate and understand their impact to more effectively inform future theory-based design with regards to coordination work within socio-technical systems. Future work may extend in many directions, including:

*Reputation system* – Related to the member browser, a reputation system could provide an additional layer of socially relevant information to collaborative editors. As Kriplean et al. (2007) pointed out, this would not be a new idea within Wikipedia. Although prior research has attempted to codify reputation within Wikipedia, early implementations of such a system (i.e., Adler & Alfaro, 2007) utilized a purely content-driven means of establishing that reputation, an approach that has proven easy to manipulate. But if a more traditional game design approach were taken to classify reputations, whereby the mechanic by which reputation was measured would be initially opaque and based on a broader spectrum of salient activity (Bogost, 2007), it may be more resilient to exploitation by individuals, as well as more capable of withstanding a supposed tyranny of the masses or unintended centralization of power.

*Role system* – As the Wikipedia Co-op and the Teahouse<sup>64</sup> have shown, there is increasing interest in ensuring a more friendly and supportive environment during the newcomer onboarding experience. Specifically, many in the community are interested in reducing the number of potential new contributors who fail to navigate the initial entry

---

<sup>62</sup> [https://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_X](https://en.wikipedia.org/wiki/Wikipedia:WikiProject_X)

<sup>63</sup> <https://en.wikipedia.org/wiki/Wikipedia:Co-op>

<sup>64</sup> <https://en.wikipedia.org/wiki/Wikipedia:Teahouse>

into becoming Wikipedian. A more codified role system, either procedurally derived or explicitly applied to, may facilitate editors' abilities to contribute in a more personally meaningful way.

Currently, there is a fair amount of overlap between the types of work that individuals complete within WikiProjects, regardless of the goals of the project itself. This work can be seen as developmental, social, motivational, or, the work of creating and improving project related content. This overlap in types of work presents an opportunity to bridge some of the structural boundaries between individual projects, to facilitate an awareness of shared efforts and abilities that could prove beneficial to disparate groups, each seeking to further individual goals in their collective ways. For instance, if multiple projects were able to access shared tools to facilitate the coordination of commonly completed tasks, the effort required to complete those tasks would be reduced, as well as the switching cost to move between projects, creating a more expansive ecosystem of collaboration. Through thoughtful intervention, it may be possible to more effectively remove these boundaries or platform constraints to allow a greater depth of interaction, to highlight new opportunities for groups to coalesce and evolve beyond the surface of the MediaWiki platform, to more fully explore the depths and potential of such a complex socio-technical system.

Development of the VTE is ongoing. It is our hope that this application, and the underlying architecture, can provide a useful entrée to future developers and designers who, in turn, may help us better understand the nature of the system itself.

## **4.8 CONCLUSION TO PART IV**

This section aims to enumerate more specifically the ways that contributors to peer production systems can effectively utilize their own motivations and creativity to contribute to those communities. By understanding the theoretical foundations of coordination within commons-based peer production communities, designers can target interventions that are likely to have a positive impact on those communities. Accepting the fact that, in a large enough population, there will exist those that are willing to contribute their own unique time and talents to a cause outside the markets and firms that traditionally have predominantly typified popular conceptions of industry and organization, it only remains to create the tools that those individuals can utilize to direct their motivations to have a positive impact on the peer production communities they reside within. Benkler (2002, p.414) supports this need for an increased capacity for task self-selection to support diverse human creativity:

*Central to my hypothesis about the information gains of peer production is the claim that human intellectual effort is highly variable and individuated. People have different innate capabilities, personal, social, and educational histories, emotional frameworks, and ongoing lived experiences. These characteristics make for immensely diverse associations with, idiosyncratic insights into, and divergent utilization of, existing information and cultural inputs at different times and in different context. Human creativity is therefore very difficult to standardize and specify in the contracts necessary for either market-cleared or hierarchically organized production. As human intellectual effort increases in importance as an input into a given production process, an*

*organization model that does not require contractual specification of effort but allows individuals to self-identify for tasks will be better at gathering and utilizing information about who should be doing what than a system that does require such specification.*

Without arguing issues of motivation or attempting to homogenize a community of contributors – contributors whom already exist with myriad motivations and an expanse of creativity among them that would be unfathomable to traverse in either an organizational manner or through technological means – I suggest that within these peer production communities, what is required is to provide tools to facilitate how community members are able to identify what they are singularly most capable of doing given their creativity. We suggest that the Virtual Team Explorer, extending directly from a theoretical examination of interactions within the community, is a step in this direction.

## Part V: Conclusion

*As peer production relies on opening up access to resources for a relatively unbounded set of agents, freeing them to define and pursue an unbounded set of projects that are the best outcome of combining a particular individual or set of individuals with a particular set of resources, this open set of agents is likely to be more productive than the same set could have been if divided into bounded sets in firms. Note that the effect changes dramatically when the resources are rival, because the value of any agent or combination of agents working on the resource is not additive to the value of any other agent or combination. In other words, the use of a rival resource excludes the use by others in a way that is not true for a purely nonrival good like information. The allocation gain is attained in allocating the scarce resource – human attention, talent, and effort – given the presence of nonrival resources to which the scarce resource is applied with only a probability of productivity.*

(Benkler, 2002, p. 422)

---

Prior sections of this dissertation have addressed the larger issue of coordination within commons-based peer production communities from a number of perspectives. In Part I, I introduced a number of prior studies, each individually focusing on coordination from a unique lens, each attempting to quantify or qualify how coordination typically *works* within Wikipedia. In the case of the primarily quantitative studies, coordination was generally interpreted as discrete actions that occur specifically within the WikiProject space. These would include the myriad WikiProject talk page or article talk page requests for coordination. More broadly, these prior studies also included under the rubric of coordination all activity that wasn't directly related to production, for instance updating templates, modules, infoboxes, or maintaining the front page of a WikiProject to ensure that it is not perceived as defunct or abandoned. My main argument in Part I of this dissertation, and my main argument running throughout the entirety of the remainder of the text, was that these operationalizations of coordination, these approximations of collaborative work in complex systems, do not provide the descriptive power to accurately identify and analyze acts of coordination as they actually occur within this peer production community. More so, I would argue that without greater granularity of observation, prior examinations have lacked the resolution necessary to draw clear conclusions about the activity occurring within those communities and have as such been unable to make clear recommendations as to the direction the community should take to ensure continued health and participation. Recalling Benkler (2002), as commons-based peer production communities can be distinguished from their predecessors, markets and

firms, by the transitory nature of those contributing to the community, any attempt to analyze that community must be from a perspective that stems explicitly from an understanding of the nature of those communities.

For example, take a traditional engineering firm in which new hires are chosen based on their prior experience and skills, where compensation is awarded in increasing amounts for tenure and tenacity, and where employee retention is resolved through, in part, reward by the employer and the social stigma or fear of unemployment by the employee. Compare this to the peer production community, in which any task can be voluntarily completed by any contributor, compensation frequently only takes the form of having a personal itch scratched<sup>65</sup>, and where the majority of new WikiProject members do little more than add their name to a participant list or add a project banner to their user page, but never return again. This type of employee entropy would likely cause traditional engineering firm to implode on itself; if each employee wrote one line of code and then quit the next day, the outlook would likely not be positive for such an organization. In peer production communities, however, it is not in spite of these attributes that they succeed but instead because of them, along with the very technology that provides the foundation for those communities – a lower barrier of entry ensures that the largest possible number of contributors can participate, increasing the modularity of work items ensures that lower barrier of entry, and providing low cost means to reintegrate those work items along with quality control mechanisms to maintain the health of the community each increase the chance that the community as a whole may thrive<sup>66</sup>.

Extending from this review of community and coordination, Part II addressed the foundations of the work that was to follow. Specifically, Part II sought to provide an initial answer to the question of how analysis and design occur within peer production communities, enumerating the ways in which both have occurred in the past while pointing towards how these tasks may be undertaken more fruitfully in the future. For instance, how can *awareness*, which has been so often cited as a factor for success throughout this dissertation, be measured in a quantitative way? How can *motivation*, reasonably present in any crowd of reasonable size, be measured, qualitatively assessed, and better understood? And through what means, tool-mediated or otherwise, do the masses engaged in peer production work actually coordinate their efforts? These studies sought to better understand not only these questions, but the manner in which they may be addressed, answered, or better understood.

Given these foundations, I suggest that one primary contribution of this dissertation is our theory of coordination for commons-based peer production communities, introduced in Part III and subsequently the central focus of the dissertation. This theory extends from the needs, identified earlier, to have a more sustainable way to both identify coordination in peer production systems as well as a more effective means of designing tools to facilitate that coordination. Further, beyond theorizing, I hoped to create a more sustainable vocabulary for addressing issues of coordination in peer production communities in the future, to facilitate a more useful common ground between analysts and designers, as well as the members of the community. Through the creation, evaluation, and analysis of this theory in the prior sections of this dissertation, I have illuminated a few primary points:

---

<sup>65</sup> From Benkler (2002) citing Eric Raymond in *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*, “Every good work of software starts by scratching a developer’s personal itch.”

<sup>66</sup> Recall that these are the three structural attributes of commons-based peer production communities from Part I of this dissertation.

- 1) The theory had the capacity to describe a broad spectrum of activities across multiple WikiProjects, and across multiple community members of varying experience, each seeking to contribute uniquely given their motivation, inclination, time, talent, and creativity.
- 2) Given that the base of the theory is grounded in the top-level categories of *identification*, *awareness*, *coordination*, and *production*, it can just as easily be utilized to analyze the activities of bots, automated, and semi-automated tools as those agents are interacting within the community by following the same rules as all human agents.
- 3) Once identified at the process level, the theory facilitated mapping coordination dependencies through time, allowing researchers to see where any individual process may succeed, falter, or fail, thus suggesting ways to more effectively identify and ameliorate those potential failures.
- 4) Beyond these individual, scenario driven processes, the theory additionally was shown to be effective at providing a far more granular approach to identifying coordinative actions across individuals, processes, and projects – in fact, the operationalizations of coordination shown in Part III could be extended to the entire community. This expansion of the theoretical representation of activity provides a more focused and surgical means of identifying relevant moments of information transfer (e.g., coordination) at multiple scopes across the community.

Ultimately, our theory of coordination for commons-based peer production communities provided a concrete means of identifying and quantifying coordination with greater resolution than prior research has achieved. This extends beyond frameworks providing rules and guidelines for successful self-organizing communities by providing the language to more effectively identify what may succeed and what fail at the individual, process, group, and community levels, in addition to providing the means by which those successes and failures can be quantitatively mapped over time.

And finally, Part IV of this dissertation provides a proof of concept of our theory of coordination, utilizing guidelines suggested by the theoretical approach to generate initial design requirements for a tool created explicitly to facilitate coordination in peer production communities – the Virtual Team Explorer. This system, in which the underlying architecture as well as the community facing components were designed to enable visibility, awareness, development, and collaboration, was then evaluated through a scenario-driven survey in which participants responded positively to our approach and to its potential capacity to improve both participation in WikiProjects as well as the function of those projects themselves.

At the outset of this dissertation I put forward three research questions that I wished to address throughout the course of the research that was to follow. Below, I will revisit each of these questions along with a discussion of the progress achieved throughout this dissertation.

**Research question 1:** *Through what means can we best analyze coordination in commons-based peer production communities to more effectively inform future research and support theory-driven design?*

As explored most prominently in Part I and Part III of this dissertation, prior interpretations of coordination have rarely positioned coordination at the level that is necessary to identify the entirety of a process-driven activity in a

peer production system. For example, without notions of simple *awareness*, the predominant means by which many Wikipedians self-select tasks to complete would be missed by any theoretical or quantitative attempts to measure that coordination – namely, the watchlist<sup>67</sup>. Given these insights into the multi-level behavior within a peer production community (i.e., individual, group, or community activity), it is more feasible to generate design requirements that can more effectively address the needs of those community members with respect to those multiple levels of interaction. More specifically, using our theory of coordination for commons-based peer production communities, design requirements can extend from the actual behavior of the community that design is intended to address, and evaluation of the design intervention can be completed with the same theory that gave rise to the emergent design requirements.

**Research question 2:** *Given the complexity of coordination and information transfer within distributed peer production communities, how can the prior interactions identified and ensuing analysis contribute to our theoretical understanding of the WikiProjects under observation?*

The granularity of information provided by our theory of coordination, introduced and explored in Part III of this dissertation, provides a novel means of comparing both group level and individual level activity between projects. Consider for instance the dependency trajectory concept maps generated from Wikipedia interactions from Part III – both maps represent data collected from actual interactions within and beyond each project, and both provide a unique lens to interpret what activity is actually being completed within those projects. While these representations are more illustrative than analytical, they provide a simple means of identifying the types of activity occurring more prominently within each WikiProject. This ability, additionally, could provide a more tractable solution to approaching many different questions regarding the community. For instance, rather than defining topic-focused projects as those whose primary focus is editing article content under the scope of a given project and task-focused projects as any not fitting the definition for topic-focused projects (Morgan et al., 2014), this approach would provide an empirical means to establish distinctions between topic or task focused WikiProjects. And as dependency trajectory concept maps likely cluster uniquely given not just the focus but the actual behavior expressed by any given WikiProject, such an empirically grounded approach to classifying project type would potentially yield multiple types of projects, beyond the topic and task focused. These results, then, could be utilized to refine the theoretical approach to collecting data, facilitating an emergent theory of coordination that would be *more effective over time* at identifying coordination activity within a given community, as well as indicating the means by which this theory could be adapted for similar use in other peer production communities – Twitter, for instance, will be explored further below.

**Research question 3:** *How can this formal, theoretical interpretation of distributed interactions inform design to best facilitate ongoing interactions and foster future growth?*

In Part IV, I demonstrate how our theory of coordination can be adopted to inform the initial design requirements of tools to facilitate coordination in peer production communities. These concrete requirements can then be referred to

---

<sup>67</sup> The use of watchlists as a predominant method of task selection was cited in (Bryant, Forte, Bruckman, 2005), as well as noted repeatedly by the Wikipedians interviewed in Part III of this dissertation.

throughout the design process to ensure that community needs are being sufficiently considered. For instance, once a unique dependency trajectory concept map has been established for any given WikiProject (see the discussion for Research question 2 above), it is possible that design interventions focusing on that project's unique signature (i.e., a coordination persona) would have a greater chance at enacting positive change among that group. As discussed above, these design requirements can then be iteratively tested and updated to ensure that tool-mediated approaches to coordination can consistently serve the needs of the communities those tools are intended to be deployed within. Future growth, potentially defined as a *group identification* dependency by our theory, can then be given proper credence during the design process and design goals can be adjusted accordingly. The VTE as described in Part IV, for instance, had as a central focus the onboarding of new contributors to existing WikiProjects. Through further testing, including both design and theory iteration, this onboarding experience could be addressed and the outcome of design interventions could be improved by adopting a lens through which to analyze and evaluate those design interventions in commons-based peer production communities.

Ultimately, I argue that our theory of coordination for commons-based peer production communities provides a grounded means of establishing a common understanding of coordination activity within peer production communities at any scale (i.e., individual, group, community). Further, given that understanding I suggest that design goals can be appropriately targeted towards the community they are intended to address. And finally, I posit that through continued iteration of both design goals and theory refinement, both design and theory can be effectively situated within the peer production community that they are intended to analyze or address.

## 5.1 FUTURE WORK

This dissertation concludes with a discussion of the future work that will be continued to further the design goals and research questions that have been laid out in the sections prior. This future work will branch along two separate but intertwined paths.

First, while the work presented above represents a substantial step towards understanding and analyzing coordination within commons-based peer production communities, I aim to further that theoretical work along several lines. First among them, I hope to continue to refine the operationalizations of *coordination* within Wikipedia and WikiProjects. As this is an emergent, iterative process, the ability to identify unique *types* of coordination among and between projects, both in the present and historically, will provide researchers with a unique lens with which to study how coordinative interactions have progressed through time. As our theory of coordination has as a central focus *awareness* and *development*, I would argue that a clear conception of this type of development would enable researchers to tell a compelling story as to the origins of the community, its continued progress, and potentially new means of explaining its success.

Related to the prior goal, I also aim to test this theory on other commons-based peer production communities. For example, if our theoretical approach to analyzing coordination can more effectively provide an analytical language to describe and explain the success of Wikipedia, I feel it would be possible (and equally engaging) to utilize such a

theoretical approach to describe the success *and failure* of both distinct WikiProjects as well as external peer production communities. For example, by utilizing our theoretical approach to collect and analyze aggregate interactions over time for currently defunct WikiProjects<sup>68</sup>, I argue a distinct trajectory of dependencies would arise, potentially suggesting reasons for WikiProject failure and, conversely, suggesting procedural or tool-mediated changes that could be adopted to increase the chances of success of other projects. With regards to other communities, I suggest that it would be equally feasible to complete a multi-scale analysis of community interactions that could, ultimately, help expand our theory of coordination's ability to describe and analyze coordination in more disparate communities. Twitter, for example, provides access to a large quantity of the data that gets channeled through its services on a daily basis. By attributing unique dependencies to specific types of directed networks, I suggest it would be possible to identify distinct types of social coordination within networks of individuals. This would, for instance, provide a novel way to facilitate bot detection or to identify and track scams over time. Similarly, I suggest that our theory of coordination could be adopted to analyze activity within Github, similar in many respects to Wikipedia but with the slightly more narrow goal of code generation. Given an analysis of coordination within this community, the capacity to understand, improve, and design for those contributors could be great, and the capacity to use that understanding to better serve *other* commons-based peer production communities would be commensurately improved as well.

Contrasting with increasing the breadth of the communities our theory of coordination is capable of describing, I also suggest increasing the granularity of the primary site of study for this dissertation. For example, while the dependency trajectory concept maps presented above have included individual, process, and project level analyses, I suggest that by limiting analysis to *type of user* a new portrait of coordination would emerge. More specifically, I suggest that by analyzing automated or semi-automated tools (i.e., bots, et al.), patterns of coordination may emerge that could suggest both what types of bots coordinate individual and project behavior most successfully, and conversely what projects may most benefit from the introduction of alternative coordinative agents.

The second area for future work I would suggest involves design progress. While the Virtual Team Explorer has been designed from conception with coordination in mind, a large scale deployment would more effectively both validate its impact on teams within Wikipedia as well as allow its use to feedback into further advancements with regard to both our theory of coordination and the tool itself. I aim to solidify the release plan for the tool and to complete development to enable its use at the scale required for deployment to all current WikiProject members and those interested in the process or the practice of the projects themselves. Ultimately, I see an approach such as the VTE as having the capacity to resolve some of the fractured practices within WikiProjects. By consolidating approaches to process, I believe that the creation and maintenance of WikiProjects can be made more transparent, and in so doing I would suggest that the means by which community members (both human and automated) interact with those projects can become more codified, structured, and mutually beneficial.

---

<sup>68</sup> For examples of defunct WikiProjects, refer to the list at [https://en.wikipedia.org/wiki/Category:Defunct\\_WikiProjects](https://en.wikipedia.org/wiki/Category:Defunct_WikiProjects), retrieved August, 2015.

In closing, the dissertation above has provided a theory of coordination for commons-based peer production communities and shown how that theory can be tested, validated, and expanded. Finally, it has shown how that theory can drive the design of future interventions into commons-based peer production communities.

# References

- Adler, B. T., & Alfaro, L. De. (2007). A Content Driven Reputation System for the Wikipedia. Proceedings of the 16th International Conference on World Wide Web, 261–270.
- Baldwin, C., & Clark, K. (2006). The architecture of participation: Does code architecture mitigate free riding in the open source development model? *Management Science*. Retrieved from <http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1060.0546>
- Beenen, G., Ling, K., Wang, X., Chang, K., Frankowski, D., Resnick, P., & Kraut, R. (2005). Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10(4), 00–00. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2005.tb00273.x/full>
- Benkler, Y. (2002). Coase's Penguin, or, Linux and The Nature of the Firm. *The Yale Law Journal*, 112(3), 369–446. Retrieved from <http://www.utexas.edu/law/journals/tlr/sources/Volume 92/Issue 2/Hemel/DSB/Hemel.fn211.01.SB.pdf>
- Benkler, Y., & Nissenbaum, H. (2006). Commons-based Peer Production and Virtue\*. *Journal of Political Philosophy*, 14(4), 394–419. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9760.2006.00235.x/full>
- Benkler, Y., Shaw, A., & Hill, B. (2013). Peer Production: A Modality of Collective Intelligence. *Mako.cc*, 1–31. Retrieved from [http://mako.cc/academic/benkler\\_shaw\\_hill-peer\\_production\\_ci.pdf](http://mako.cc/academic/benkler_shaw_hill-peer_production_ci.pdf)
- Beschastnikh, I., Kriplean, T., & McDonald, D. W. (2008). Wikipedian self-governance in action: Motivating the policy lens. *Proc. ICWSM 2008*, 9. Retrieved from [www.aaii.org/Papers/ICWSM/2008/ICWSM08-011.pdf](http://www.aaii.org/Papers/ICWSM/2008/ICWSM08-011.pdf)
- Birnholtz, J., & Ibara, S. (2012). Tracking changes in collaborative writing: edits, visibility and group maintenance. *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 809–818). New York, NY, USA: ACM. doi:10.1145/2145204.2145325
- Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., & Hellmann, S. (2009). DBpedia - A crystallization point for the Web of Data. *Journal of Web Semantics*, 7, 154–165. doi:10.1016/j.websem.2009.07.002
- Bogost, I. (2007). *Persuasive games: The expressive power of videogames*. MIT Press.
- Bryant, S. L., Forte, A., & Bruckman, A. (2005). Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work (GROUP'05)*, 1–10. Retrieved from <http://dl.acm.org/citation.cfm?id=1099205>
- Callender, C., & Cohen, J. (2006). There Is No Special Problem About Scientific Representation. *Theoria*, 4548, 7–25.
- Calvino, I. (1974). *Invisible Cities*. 1972. Trans. William Weaver. New York: Harcourt Brace Jovanovich.
- Charmaz, K. (2014). *Constructing grounded theory*. Sage.
- Chi, E., & Suh, B. (2008). Providing social transparency through visualizations in Wikipedia. *Social Data Analysis Workshop at CHI 2008*, 1–4. Retrieved from [cs.smith.edu/classwiki/images/a/a3/Social\\_transparency\\_visualization\\_wikipedia.pdf](http://cs.smith.edu/classwiki/images/a/a3/Social_transparency_visualization_wikipedia.pdf)
- Cohen, J. A. (1960). Coefficient of agreement for nominal scales', *Educational and Psychological Measurement*, 20, 37–46. *Educational and Psychological Measurement*.

- Cosley, D., Frankowski, D., Terveen, L., & Riedl, J. (2007). SuggestBot: using intelligent task routing to help people find work in wikipedia. *IUI '07 Proceedings of the 12th International Conference on Intelligent User Interfaces*. Retrieved from [dl.acm.org/citation.cfm?id=1216309](http://dl.acm.org/citation.cfm?id=1216309)
- Cosley, D., Frankowski, D., & Kiesler, S. (2005). How oversight improves member-maintained communities. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '05 Human Factors in Computing Systems*, 11. doi:10.1145/1054972.1054975
- Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*. Retrieved from <http://www.springerlink.com/index/j8pw5067pu421747.pdf>
- Crowston, K., Li, Q., Wei, K., Eseryel, U. Y., & Howison, J. (2007). Self-organization of teams for free/libre open source software development. *Information and Software Technology*, 49(6), 564–575. doi:10.1016/j.infsof.2007.02.004
- Crowston, K. (2003). A taxonomy of organizational dependencies and coordination mechanisms. In T. Malone, K. Crowston, & G. Herman (Eds.), *Organizing Business Knowledge: The MIT Process Handbook* (pp. 85–108). Cambridge, MA: MIT Press. Retrieved from <http://ccs.mit.edu/papers/CCSWP174.html>
- Crowston, K. (1991). *Towards a Coordination Cookbook: Recipes for Multi-Agent Action*. (Doctoral dissertation). Retrieved from <http://lib.washington.edu>. (OCLC No. 24343522).
- Dourish, P. (2003). The appropriation of interactive technologies: Some lessons from placeless documents. *Computer Supported Cooperative Work: CSCW: An International Journal*, 12(4), 465–490. doi:10.1023/A:1026149119426
- Dourish, P., & Bellotti, V. (1992). Awareness and coordination in shared workspaces. *Proceedings of the 1992 ACM Conference on Computer-supported cooperative work*, (November), 107–114. Retrieved from <http://dl.acm.org/citation.cfm?id=143468>
- Elgin, C. Z. (2008). Exemplification, Idealization, and Scientific Understanding, Ch. 5. In M. Suarez (Ed.), *Fictions in Science: Philosophical Essays on Modeling and Idealization* (1st ed., pp. 77–90). New York: Routledge.
- Engeström, Y. Activity theory and individual and social transformation. In Engeström, Y., Miettinen, R. and Punamaki, R.L., editors, *Perspectives on Activity Theory*, Cambridge University Press: Cambridge, UK, 1999, 19-38
- Farzan, R., Dabbish, L., Kraut, R. E., & Postmes, T. (2011). Increasing Commitment to Online Communities by Designing for Social Presence. *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work - CSCW '11*, 321–330. doi:10.1145/1958824.1958874
- Forte, A., Kittur, N., Larco, V., Zhu, H., Bruckman, A., & Kraut, R. E. (2012). Coordination and beyond: social functions of groups in open content production. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 1–10. Retrieved from [dl.acm.org/citation.cfm?id=2145270](http://dl.acm.org/citation.cfm?id=2145270)
- Forte, Andrea, Vanessa Larco and Amy Bruckman. (2009) Decentralization in Wikipedia Governance. *Journal of Management Information Systems*. 26(1) pp 49-72
- Forte, A., & Bruckman, A. (2008). Scaling consensus: Increasing decentralization in Wikipedia governance. *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual Conference*, 15–17. Retrieved from [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=4438860](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4438860)
- Geiger, R. S. (2014). Bots, bespoke, code and the materiality of software platforms. *Information, Communication & Society*, 1–15. doi:10.1080/1369118X.2013.873069

- Geiger, R., & Halfaker, A. (2013a). Using edit sessions to measure participation in Wikipedia. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 861–869. Retrieved from <http://dl.acm.org/citation.cfm?id=2441873>
- Geiger, R., & Halfaker, A. (2013b). When the levee breaks: without bots, what happens to Wikipedia's quality control processes? *WikiSym '13 Proceedings of the 9th International Symposium on Wikis and Open Collaboration*, 0–5. Retrieved from <http://dl.acm.org/citation.cfm?id=2491061>
- Geiger, R., & Ribes, D. (2011). Trace Ethnography: Following Coordination through Documentary Practices. *System Sciences (HICSS)*, 2011 44th, 1–10. doi:10.1109/HICSS.2011.455
- Gilbert, E. (2012). Designing social translucence over social networks. *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI '12*, 2731. doi:10.1145/2207676.2208670
- Gilbert, M., & Zachry, M. (2015). Towards a Class-Based Model of Information Organization in Wikipedia. *HCI International 2015 Lecture Notes in Computer Science (LNCS) series*.
- Gilbert, M., Morgan, J., McDonald, D. W., & Zachry, M. (2013). Managing complexity: strategies for group awareness and coordinated action in Wikipedia. *Proceedings of the 9th International Symposium on Open Collaboration*, 1–8. Retrieved from <http://dl.acm.org/citation.cfm?id=2491060>
- Gutwin, C., Penner, R., & Schneider, K. (2004). Group awareness in distributed software development. *Proceedings of the 2004 ACM conference on Computer supported cooperative work - CSCW '04*, 72. doi:10.1145/1031607.1031621
- Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2013). The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline. *American Behavioral Scientist*, 57(5), 664–688. doi:10.1177/0002764212469365
- Halfaker, A., Kittur, A., & Riedl, J. (2011, October). Don't bite the newbies: how reverts affect the quantity and quality of Wikipedia work. In *Proceedings of the 7th international symposium on wikis and open collaboration* (pp. 163-172). ACM.
- Haley, S. M., & Osberg, J. S. (1989). Kappa coefficient calculation using multiple ratings per subject: a special communication. *Physical Therapy*, 69(11), 970–974.
- Heilman, J. M., Kemmann, E., Bonert, M., Chatterjee, A., Ragar, B., Beards, G. M., ... & Laurent, M. R. (2011). Wikipedia: a key tool for global public health promotion. *Journal of medical Internet research*, 13(1).
- Monroy-Hernández, a. (2013). The new war correspondents: The rise of civic media curation in urban warfare. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 1443–1452. doi:10.1145/2441776.2441938
- Hill, B.M. (2013). *Essays on Volunteer Mobilization in Peer Production* (Doctoral dissertation). Retrieved from <http://hdl.handle.net/1721.1/86240>.
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–88. doi:10.1177/1049732305276687
- Karau, S., & Williams, K. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, 65(4), 681–706. Retrieved from <http://psycnet.apa.org/journals/psp/65/4/681/>

- Kittur, A., Pendleton, B., & Kraut, R. E. (2009). Herding the cats: the influence of groups in coordinating peer production. *WikiSym '09 Proceedings of the 5th International Symposium on Wikis and Open Collaboration*. Retrieved from <http://dl.acm.org/citation.cfm?id=1641321>
- Kittur, A., & Kraut, R. (2008). Harnessing the wisdom of crowds in wikipedia: quality through coordination. *Proceedings of the 2008 ACM Conference on ...*. Retrieved from <http://dl.acm.org/citation.cfm?id=1460572>
- Kittur, A., Chi, E., Pendleton, B. A., Suh, B., & Mytkowicz, T. (2007). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *World Wide Web*, 1(2), 19. Retrieved from <http://edouard-lopez.com/fac/ICPS - S7/Complexity/2008-Wikipedia-As-A-Complex-System/Power of the Few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie.pdf>
- Kittur, A., Suh, B., Pendleton, B. A., & Chi, E. H. (2007). He says, she says: conflict and coordination in Wikipedia. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 453–462). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1240698>
- Kraut, R. E. & Resnick, P. (2012). *Building successful online communities: Evidence-based social design*. Cambridge, MA: MIT Press
- Krieger, M., Stark, E., & Klemmer, S. (2009). Coordinating tasks on the commons: designing for personal goals, expertise and serendipity. *CHI '09 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1485–1494. Retrieved from <http://dl.acm.org/citation.cfm?id=1518927>
- Kriplean, T., Morgan, J., & Freelon, D. (2011). ConsiderIt: Improving structured public deliberation. *Proceedings of the CHI '11 Extended Abstracts on Human Factors in Computing Systems*, 1831–1836. Retrieved from <http://dl.acm.org/citation.cfm?id=1979869>
- Kriplean, T., Beschastnikh, I., McDonald, D. W., & Golder, S. A. (2007). Community, consensus, coercion, control: CS\*W or How Policy Mediates Mass Participation. *Proceedings of the 2007 International ACM Conference on Conference on Supporting Group Work - GROUP '07*, 167–176. doi:10.1145/1316624.1316648
- Landis, J. R., & Koch, G. G. (1977a). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. doi:10.2307/2529310
- Lange, D., Böhm, C., & Naumann, F. (2010). Extracting structured information from Wikipedia articles to populate infoboxes. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM '10)*, 1661–1664. doi:10.1145/1871437.1871698
- Latour, Bruno. (1996). *Aramis, or, The love of technology*. Vol. 1996. Cambridge, MA: Harvard University Press.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press: Cambridge, UK.
- Malone, T. W., & Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Computing Surveys*, 26(1), 87–119. doi:10.1145/174666.174668
- Malone, T. W. (1987). Modeling coordination in organizations and markets. *Management science*, 33(10), 1317–1332.
- Marwell, G., Oliver, P. (1993). *The Critical Mass in Collective Action*. Cambridge University Press.
- McDonald, D. W., Gokhman, S., & Zachry, M. (2012). Building for Social Translucence : A Domain Analysis and Prototype System. *Proceedings of CSCW*, 637–646. doi:<http://doi.acm.org/10.1145/2145204.2145301>

- McGrath, J. E. (1991). Time, interaction, and performance (TIP) A Theory of Groups. *Small Group Research*, 22(2), 147–174. Retrieved from [sgr.sagepub.com/content/22/2/147.short](http://sgr.sagepub.com/content/22/2/147.short)
- Monge, P. R., & Contractor, N. S. (2003). *Theories of communication networks*. Oxford University Press.
- Morgan, J., Gilbert, M., McDonald, D. W., & Zachry, M. (2014). Editing beyond articles: diversity & dynamics of teamwork in open collaborations. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 550–563. Retrieved from <http://dl.acm.org/citation.cfm?id=2531654>
- Morgan, J., Gilbert, M., McDonald, D. W., & Zachry, M. (2013). Project talk: coordination work and group membership in WikiProjects. *Proceedings of the 9th International Symposium on Open Collaboration*. Retrieved from <http://dl.acm.org/citation.cfm?id=2491058>
- Morgan, J., Bouterse, S., Walls, H., & Stierch, S. (2013). Tea and sympathy: crafting positive new user experiences on wikipedia. *CSCW '13 Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 839–848. Retrieved from <http://dl.acm.org/citation.cfm?id=2441871>
- Nardi, B., Whittaker, S., & Schwarz, H. (2002). NetWORKers and their activity in intensional networks. *Computer Supported Cooperative Work* 11, 205–242. Retrieved from <http://www.springerlink.com/index/BJ722AJU32Q7D4FL.pdf>
- Ostrom, E. (2000). Collective action and the evolution of social norms. *The Journal of Economic Perspectives*, 14(3), 137–158. Retrieved from <http://www.jstor.org/stable/2646923>
- Pace, S. (2004). A grounded theory of the flow experiences of Web users. *International Journal of Human-Computer Studies*, 60(3), 327–363. doi:10.1016/j.ijhcs.2003.08.005
- Priedhorsky, R., Chen, J., Lam, S. (Tony) K., Panciera, K., Terveen, L., & Riedl, J. (2007). Creating, destroying, and restoring value in wikipedia. *Proceedings of the 2007 International ACM Conference on Conference on Supporting Group Work - GROUP '07*, 259. doi:10.1145/1316624.1316663
- Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying Common Identity and Bond Theory to Design of Online Communities. *Organization Studies*, 28(3), 377–408. doi:10.1177/0170840607076007
- Ribes, D., Jackson, S., Geiger, S., Burton, M., & Finholt, T. (2013). Artifacts that organ-ize: Delegation in the distributed organization. *Information and Organization*, 23(1), 1–14. doi:10.1016/j.infoandorg.2012.08.001
- Robinson, M. (2004). *Housekeeping: A Novel*. Macmillan.
- Rorty, R. (Ed.). (1992). *The linguistic turn: Essays in philosophical method*. University of Chicago Press.
- Rughinis, C., & Matei, S. (2013). Learning through Massively Co-Authored Biographies: Making Sense of Steve Jobs on Wikipedia through Delegated Voice. *2013 19th International Conference on Control Systems and Computer Science*, 590–596. doi:10.1109/CSCS.2013.62
- Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company. ISBN:0-395-61556-9
- Shaw, A., & Hill, B. M. (2014). Laboratories of Oligarchy? How the Iron Law Extends to Peer Production. *Journal of Communication*, 1–24. doi:10.1111/jcom.12082
- Shirky, C. (2010). *Cognitive surplus: How technology makes consumers into collaborators*. Penguin.

- Skeels, M., & Grudin, J. (2009). When social networks cross boundaries: a case study of workplace use of facebook and linkedin. Proceedings of the ACM 2009 international conference on Supporting group work. Retrieved from <http://dl.acm.org/citation.cfm?id=1531689>
- Suh, B., Convertino, G., Chi, E. H., & Pirolli, P. (2009, October). The singularity is not near: slowing growth of Wikipedia. In Proceedings of the 5th International Symposium on Wikis and Open Collaboration (p. 8). ACM.
- Social Proof (n.d.). Wikipedia. Retrieved from [https://en.wikipedia.org/wiki/Social\\_proof](https://en.wikipedia.org/wiki/Social_proof).
- Sproull, L., Conley, C., & Moon, J. (2005). Prosocial behavior on the net. The social net: Understanding human behavior, 139–162. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.114.809&rep=rep1&type=pdf>
- Viégas, F. B., Wattenberg, M., Kriss, J., & Van Ham, F. (2007). Talk before you type: Coordination in Wikipedia. In System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on (pp. 78–78). IEEE. Retrieved from [ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=4076527](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4076527)
- Viégas, F., Wattenberg, M., & McKeon, M. (2007). The hidden order of Wikipedia. Online Communities and Social Computing, 0–9. Retrieved from [http://link.springer.com/chapter/10.1007/978-3-540-73257-0\\_49](http://link.springer.com/chapter/10.1007/978-3-540-73257-0_49)
- Viégas, F. B., Wattenberg, M., & Dave, K. (2004). Studying cooperation and conflict between authors with history flow visualizations. Proceedings of the SIGCHI conference on Human factors in computing systems (Vol. 6, pp. 575–582). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=985765>
- Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the Process of Sensemaking. Organization Science, 16(4), 409–421. doi:10.1287/orsc.1050.0133
- Wenger, E. (2000). Communities of practice and social learning systems. Organization. Retrieved from <http://org.sagepub.com/content/7/2/225.short>
- Wu, F., & Weld, D. S. (2010). Open information extraction using Wikipedia. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL '10), (July), 118–127. Retrieved from <http://dl.acm.org/citation.cfm?id=1858681.1858694>
- Wu, F., & Weld, D. S. (2008). Automatically Refining the Wikipedia Infobox Ontology. Proceeding of the 17th International Conference on World Wide Web WWW 08, (January), 635–644. doi:10.1145/1367497.1367583
- Zhang, P, and Galletta, D. eds. (2006). Human-Computer Interaction and Management Information Systems: Foundations. Armonk, New York: M.E. Sharpe, Inc.
- Zhao, D., & Rosson, M. (2009). How and why people Twitter: the role that micro-blogging plays in informal communication at work. Proceedings of the ACM 2009 international conference on Supporting group work (GROUP'09), 243–252. Retrieved from <http://dl.acm.org/citation.cfm?id=1531710>
- Zhu, H., Kraut, R., & Kittur, A. (2012). Organizing without formal organization: group identification, goal setting and social modeling in directing online production. Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. Retrieved from <http://dl.acm.org/citation.cfm?id=2145204.2145344>

## **About the Author**

Michael Gilbert was born in Seattle, Washington, lived in Spokane, raised in Arizona, moved out of Friday Harbor, and completed this dissertation while living on a sailboat on Lake Union, back in Seattle. Prior to completing the Human Centered Design & Engineering PhD program at the University of Washington, he received his Master of Science in Information Management and his Bachelor of Arts in English with a creative writing focus, both also at the University of Washington. Finally, in 2016, after a lengthy academic career and a number of years spent as a programmer in the video game industry, he embarked on a new adventure as a User Experience Researcher for Google.

# Appendices

## APPENDIX A: COORDINATION CODESHEET DESCRIPTION FOR DIRECTED CONTENT ANALYSIS

What follows is the earliest incarnation of the codesheet utilized to classify moments of information transfer within interview transcripts during the directed content analysis process, prior to selective coding with our own coordination framework.

---

**Malone and Crowston’s common dependencies and example alternative coordination mechanisms for dealing with them. Originally from (Malone & Crowston, 1994).**

<i>Dependency</i>	<i>Alternative coordination mechanism</i>
<i>Shared resources</i>	“First come, first serve,” priority order, budgets, managerial decision
Task assignments	(same as for “Shared resources”)
<i>Producer/Consumer relationships</i>	
Prerequisite constraints	Notification, sequencing, tracking
Transfer	Inventory management (e.g. “just in time,” “Economic order quantity”)
Usability	Standardization, ask users, participatory design
Design for manufacturability	Concurrent engineering
<i>Simultaneity constraints</i>	Scheduling, synchronization
<i>Task/Subtask</i>	Goal selection, task decomposition

### Coding Cheat Sheet -

What follows is a more detailed explanation of each of the dependencies identified in the original Coordination Theory taxonomy. While coding, if there is any ambiguity as to which dependency should be categorized with any given interaction (the highlighted portions of the text), consider the possibility that a new category should be added to expand the current taxonomy. And, as mentioned earlier, it is also possible that any given interaction be categorized with more than a single dependency.

*A couple points on identifying and distinguishing between dependencies:*

When you are looking at an interaction, it may be helpful to ask yourself a few simple questions:

- What is the scope of analysis? Are we talking from the individual’s perspective or the group’s perspective, from the perspective of a potential newbie needing help or the trainer that will be helping the newbie?
- What is the action that’s taking place?
- What is the desired result of that action?

Given these questions, you should be able to narrow down which dependency may fit. If none of these dependencies fit after reflecting on these questions, mark the action as “Other” and add your Initial Code. I’m adding examples below as I find them; if you find anything that appears to be an especially good fit with any of the dependencies feel free to add it in the text or as a comment.

***Shared resources:Other:***

Whenever more than one activity depends on a resource, a shared resource dependency may exist. In terms of human actors, this may take the form of two or more individuals requiring a single tool at the same time. An offline example would be two cooks in a kitchen both requiring the same knife at the same time. Within our target system, an example may be two editors attempting to edit the same article at the same time, thereby creating potential conflicts (although this specific challenge has not yet been seen in our data collection). Computationally, one example would be threading in a modern processor, whereby jobs are queued by both time and priority so that computation can proceed in a largely coordinated fashion. Potential means of mediating this dependency include queuing, market-based transaction models where the resource (including time, money, human or non-human agents) can be bid upon by actors, managerial decision, or popular vote.

**Example:**

*02\_0075 A: I suppose I could say that the more eyes on the talk page, the faster the response.*

I would include this as Shared Resources:Other, as the statement is implying that a greater number of participants equates to a greater chance of receiving a speedy response. In other words, the action is to post a question, the desired result is to get a fast response, the dependency identified is Shared Resources, and one mechanism for resolving that dependency is to have a greater number of people watching that page to ensure that fast response. Conversely, if this dependency is unmet (ie, not enough people watching the page), a response might not be received. It’s important to note the scope of analysis for this example - this is an observation at the level of the project. By carefully looking at the context of the statement, we can see that the interviewee was talking at a higher, more abstract level (“the more eyes on the talk page, the faster the response”), indicating that the dependency identified should be at that similar level. If he had, for instance said “I posted a question on the talk page and it took forever to get a response” it would have been more in line with Producer/consumer relationships.

*02\_0015 A: I worked a lot with WikiProject Open Access, which actually had less collaboration, but the collaborators that it did attract put in a lot of work.*

I would also include statements about project selection as a Shared resources dependency. In this example the scope of analysis is on the individual (“I worked a lot...”), the action and the desired result is the work completed or the notion of group alignment (ie, no work necessarily has to have taken place), the dependency is Shared Resources:Other, and one potential mechanism to resolve that dependency is to successfully “join” the group (self-identification as a group member). This, though, assumes that the scope of analysis is the individual in that the shared resource is the finite amount of that individual’s time. From the perspective of the project, the desired result

may be to increase membership numbers to more successfully complete project related tasks. In this case the dependency would still be Shared resources, but the resource in question would then be the project members (and not individual time). Statements about recruitment or project success equating to greater membership may fall under this category.

It's also important to note the distinction between selecting an article to work on, which may be an example of Shared Resources:Other, as it's the micro version of selecting a project to join, and selecting an article to work on by subject area, which is more specific. That specificity implies that the act of selecting is no longer a generic statement of "what should I do with my time," where time is the resource, and instead is a statement of directed intent. The latter, which I would classify as Other, is motivated by a specific interest to improve a specific area and would not fall under the current Coordination Theory taxonomy. Further, if the statement were selecting an article to work on by subject area to inform a specific reader(s), the dependency would actually be Producer/consumer relationship, as in this last statement there is a specific actor the information is directed at.

#### ***Shared Resources:Task assignment:***

As this is a subset of shared resources, most examples and mediating coordination mechanisms for that dependency also apply here. The distinguishing feature of task assignment is the focus on the scarcity of time of the actors that will be responsible for completing an action. Again, this could include human, non-human, or information processing components within the larger system. Within Wikipedia, an example may be in the allocation of editor efforts to improve article content, to enable bots to update project pages, or from the perspective of tools, to base a decision to enable a bot for a specific project based on the properties of that project. More specifically, the Hot Articles bot is a tool that projects can enable to list the most actively edited articles under the scope of that project within recent history (configurable to either the last three or last seven days). If the number of articles under the scope of the project are too great, however, the time required to process the recent editing data is too great so the subscription request can not be honored.

#### ***Producer/consumer relationships:Other:***

This dependency occurs when one activity produces something (an output) that is required by an ensuing activity (an input), implying a temporal process involving multiply chained activities. This is a fairly common dependency given the scope of analysis for any given activity. For example, to improve an article at the highest scope may just be a matter of adding quality content. But at a more granular level, this may also entail posting to the project talk page to request assistance with article editing, posting to the article talk page to ask for clarification about article details, or simply finding the time to sit down and make the changes that are required. In this more granular illustration, each of the prior examples represents a producer/consumer relationship, a producer/consumer relationship, and a shared resource dependency, respectively. For our purposes we will aim to deconstruct tasks as much as possible given the interview responses to most accurately categorize the activities described by our interviewees.

***Producer/consumer relationships: Prerequisite constraints:***

This is a common type of producer/consumer dependency, indicating a series of tasks that must be completed in order for a goal to be achieved. The dependency arises when there must be some sort of notification between each task, so that at the completion of the first, the second can begin. Notification can be enacted through human or non-human agents, and can take the form of actively pushing information to predefined locations (for instance, updating a task list with the current state of an activity's progress) or through passive awareness (for instance, updating an article to achieve good article status but expecting other editors to promote it based on its merits). Each of these approaches have their own costs associated. In the first example, coordination costs would be higher due to the additional activity required by the editor to update the task list, and the additional activity of other editors to monitor that list, and in the second example in the knowledge dependency required for additional editors to realize that the article should now be promoted.

**Example:**

02\_0019 A: [So, if there's something that happens on certain WikiProjects, I'll see the message and then I might respond to it.](#)

In this statement, we can see that the interviewee is discussing an interaction from a personal perspective (as opposed to the higher level or more abstract hypothetical perspective, which may have made this example be closer to a Shared Resources dependency). In this case, the action and the desired result is the response, the dependency is Producer/consumer relationships: Prerequisite constraints, and one potential mechanism for resolving that dependency is to have a watchlist that informs you when a question has been asked that may require a response. This is unique from the Shared Resources example above in that, without notification such as a watchlist or email alert, no action would be taken. It is also unique from Shared Resources: Task Assignment since nothing is being assigned, the interviewee is self-selecting tasks based on awareness of ongoing actions within Wikipedia.

***Producer/consumer relationships: Transfer:***

In this instance, the activity that produces an object typically has to physically transfer that object to the consumer. This dependency would be more frequently seen in manufacturing, where, for instance, if a car is being built the frame must be delivered before the doors can be put on. In the case of our target system, physical transfer may not be as common but it still may be observed in some instances. One example may be a researcher requiring a dump of the entire history of Wikipedia before continuing work, necessitating waiting until the dump is either downloaded (transfer of bits) or delivered (transfer of physical media). In either case this dependency would additionally involve the required storage of the object received, either through physical means (for instance, the back storage room at a supermarket containing food waiting to be placed on the shelves) or through informational means (the disk space required to store the dump of Wikipedia data).

***Producer/consumer relationships:Usability:***

Potentially more familiar to those in the human-centered design field is the usability dependency. This dependency refers to the need for any object created by the producer to be delivered to the consumer in a form that is usable. In the case of our target system, this may take the form of a question being asked on an article talk page that, if an answer in response to that question does not actually address the needs of the asker, the answer will not be usable. Potential means of ameliorating this dependency include standardization, potentially through tool-mediated means of ensuring that the outputs of one activity match the inputs required by the next, or participatory design, through actively involving stakeholders in a design or creative process to ensure that the object delivered is as expected and required.

**Example:**

*02\_0300 A: Another thing that has been proposed is to improve the quality of the help guide's tutorials, that way that a person acting alone will be able to go to Wikipedia, not interact with other humans, but somehow come to understand how it is that they can do whatever it is they want to do on Wikipedia.*

In this statement, we can see that the scope of analysis is from the perspective of an individual as opposed to a WikiProject, indicating that what we have is a Producer/consumer relationships:Usability dependency. If this were from the perspective of the project, it would likely be a Shared resources dependency (for example, if the interviewee suggested that projects would have greater chances of retaining members and completing their work if there were better tutorials). So, for this example, the actor is a newcomer to Wikipedia, the action is to read the online tutorials, the desired result is to become competent enough to “do whatever it is they want to do” on Wikipedia, the dependency is Producer/consumer relationships:Usability, and one potential mechanism to resolve that dependency is to ensure online help resources are of sufficient quality to effectively train new Wikipedians.

*02\_0301 A: So some examples of that are, say, reform of the navigation system on the main page of Wikipedia, or redesign of the help archives, also theres something called the Wikipedia Adventure, which is an online tutorial that guides new users with a special interface on how to edit Wikipedia, so they do fake edits in a controlled environment instead of live edits in Wikipedia, and they do this with a computer system helper, something like Microsoft Clippy, that kind of thing.*

This statement is actually a continuation of the sentence above. I would include this as an example of Producer/consumer relationships:Usability primarily from the first part of the sentence - “So some examples of that are, say, reform of the navigation system on the main page of Wikipedia, or redesign of the help archives.” This statement seems to indicate that it is the structure of Wikipedia that needs to be improved for a new Wikipedian to be able to effectively navigate within the community. In this instance, the action is simply to be able to find help in the archives or navigate the main page, the desired result is to find the correct resources or effective navigation, the dependency is Producer/consumer relationships:Usability, and one potential mechanism for resolving this dependency is to improve the site structure to ensure that it is navigable and help is findable.

The second part of this sentence is slightly more ambiguous - "... also, there's something called the Wikipedia Adventure, which is an online tutorial that guides new users with a special interface on how to edit Wikipedia, so they do fake edits in a controlled environment instead of live edits in Wikipedia, and they do this with a computer system helper, something like Microsoft Clippy, that kind of thing." From the perspective of an individual receiving training, the action would be playing the Wikipedia Adventure game, the desired result would be proficiency editing Wikipedia, and the dependency would be a lack of sufficient knowledge to successfully contribute. This doesn't involve any transfer of information (which would imply that the dependency is Prerequisite constraints), but a required increase in knowledge. This notion of necessary development over time is not currently included in the Coordination Theory taxonomy, and as such this would be marked Other. Alternatively, from the perspective of an individual delivering training (which, given the context of the sentence above seems like it might be more accurate), the action would be training, the desired result would be to create more experienced Wikipedians, and, in the case of the Wikipedia Adventure game, the dependency would likely be Producer/consumer relationship:Usability. In other words, if we consider the failed dependency (a user played the game but still did not become a proficient Wikipedian because the game was poorly designed), it becomes more clear that the failure was related to a Usability problem with the game.

***Producer/consumer relationships:Design for manufacturability:***

Related to usability, this dependency also indicates a need for the outputs of one activity to be suitable for the inputs of another. Frequently, this dependency can be mediated through a concurrent design process, whereby stakeholders for a consuming activity work in conjunction with the stakeholders for the producing activity to ensure that the object or information delivered is suitable for the goals of the larger organization.

***Simultaneity constraints:***

This dependency indicates a situation in which multiple tasks need to occur at the same time, or conversely, a situation in which multiple tasks can not occur at the same time. Similar to the shared resources example above, within our target system this may occur when two editors attempt to update the same article at the same time, thereby creating a conflict (in this instance the article is being referred to as both a resource and a task, a distinction that can be made more clear if necessary). One potential coordination mechanism that could mediate this dependency could be synchronization, through either human or tool-mediated means, to ensure that task management is completed in a manner that does not cause this shared resource conflict.

***Task/subtask:***

As mentioned above, depending on the scope of analysis many tasks can be broken down into smaller constituent parts. For our purposes this can be broken down into two parts. First, the overall goal of a task must be chosen, a

process referred to as goal selection. After, the subtasks required to attain that goal must be selected, a process referred to as goal decomposition. In practice, our goal will be to break the selected goal down into its component subtasks inasmuch as possible given the description of that process by the interviewees. Ultimately, this dependency is one of structure, requiring actors to sequentially or simultaneously identify (goal selection) and complete (goal decomposition) the constituent components of a larger problem to achieve a desired effect. In our target system, this may be seen in the manner by which an editor may choose to encourage newcomers to join a project, potentially involving many individual approaches to both entice and retain those editors as contributing members of the project.

## **APPENDIX B: COORDINATION CODESHEET DESCRIPTION FROM GROUNDED THEORY ANALYSIS**

The following sections describe the current incarnation of our focused codes. These were arrived at through an iterative process of coding and validating interview transcript data during a series of Directed Research Groups at the University of Washington<sup>69</sup>. The following codesheet comprises the most recent incarnation of that codesheet, which was utilized by each member of the Directed Research Group to classify interactions identified within interview transcripts by the author and subsequently evaluated using Cohen's Kappa.

### **Group identification**

#### **Project selection, project knowledge, outreach, recruitment**

In traditional top-down organizations, a new member typically joins a team when they are hired. Individual roles and responsibilities within that team are dictated by management, and continued involvement with that team is typically motivated by remuneration (ie, money) and contingent on each individual's ability to successfully complete the tasks assigned to them. In Commons-Based Peer Production communities where contributions are typically made by unpaid volunteers, there is no notion of "hiring." In these communities, involvement is self-directed. As a result, there is typically very little or no hierarchy in these organizations, with work organized in a bottom-up fashion given the whims and availability of those contributing their efforts to the goals of the team. This dependency, if universally unmet, implies the dissolution of the group. Potential means of resolving this dependency include outreach campaigns, common bond (ie, increase the bond between individual group members through improved means of communication or defined collaborations), or common identity (ie, increase an individual's bond with the group by, for instance, advertising one's own continued involvement with that group, such as T-shirts with a brand on them, or creating an online profile which links back to the projects one is a member of). As such, this dependency includes both the self-identification with a group (ie, I would like to join that group), as well as outreach and recruitment efforts on behalf of the group to increase cohesion and participation (ie, You should come join this group), each depending on the scope of analysis (acting out of personal interest versus acting on behalf of the group).

### **Role identification**

#### **Task selection, task identification, role selection**

Similar to Group identification, Role identification refers to group members capacity to (or need to) identify how they want to contribute to a group, and to select tasks appropriate to that role. To be more specific, this would include both task selection and role selection, in that it includes both the work a member self-selects to complete and the role the member chooses, potentially constraining future tasks to self-select (ie, taking a leadership role will lead to more leadership tasks and requests - in a volunteer organization, there is not really a concept of "Task assignment"). Again, in contrast to traditional organizations where task assignment is typically top-down, in CBPP

---

<sup>69</sup> Directed Research Groups provide an opportunity for University of Washington undergraduate and graduate students to participate in ongoing research projects by other students and faculty. More details on these groups can be found at <http://www.hcde.washington.edu/research/directed>.

communities tasks are more often voluntary and self-selected. Further, the level of involvement as well as the individual role can (and often does) change rapidly, with personal interest, time available, organizational knowledge (ie, do you know the movers and shakers, do you know the structure, culture, and goals of the group?), and subject area expertise all playing a part in the types of work an individual may self-select to complete for a group. Potential means of resolving this dependency include those suggested to resolve Group identification as well as task identification (ie, provide tools to ensure that those looking to contribute can find a valuable way to do so), task assignment (ie, actually tell new group members the tasks they need to complete), or mentorship (ie, have explicitly defined mentorship roles which help new group members develop the skills necessary to valuably contribute to the group's goals). This would include my focused category of both "Fostering growth" and "Humanizing project members," both of which are more examples of coordination mechanisms than dependencies.

## **Group awareness**

**Monitoring (at the generic level), monitoring watchlists, monitoring vandalism. Awareness of content**

In an organization with little or no explicit hierarchy, with ill-defined roles and constantly shifting membership, awareness of the context and culture of that group becomes far more important, especially for new group members, to ensure that individual contributions are aligned with the needs and culture of the group. This top-level category will include monitoring the pages under the scope of a given project, monitoring for vandalism, and generally maintaining awareness of the content the project claims responsibility for. For the more specific goal of maintaining awareness of project members, project tasks, or the traces more directly responsible for project coordination, the more detailed awareness categories below should likely be considered.

## **Cultural awareness**

**Knowledge of how interactions should occur (at an interpersonal level), knowledge of shared language, knowledge of group norms, knowledge of policy and practice (i.e., "resistance to change"), identifying emotional states**

Cultural awareness refers to an individual's ability to effectively construct knowledge regarding effective means of interacting within a group. While Structural awareness refers to the actual construction of a space that lends itself to increasing knowledge about how to operate in that space, cultural awareness is instead referring to knowing how the nature of individual interactions will facilitate effective interactions within that space. For instance, if debates within an online group are typically resolved through discourse (ie, informed debate as opposed to vote), an individual would need to be familiar with the practice of communication within that group to successfully participate in that debate. Or, if a group is known for its tight adherence to the rules (or conversely, its laid back approach to contribution), interactions within that group that go against those culturally adopted values would likely be detrimental to one's continued successful involvement. Whereas Structural awareness refers to an ability to successfully construct a mental model of the physical space to effectively interact within that space, Cultural awareness refers to the ability to interact in that space over time in a manner that is in alignment with group norms.

## **Social awareness**

### **Knowledge of others' group roles, knowledge to navigate social structure**

Social awareness refers to an individual's capacity to construct knowledge about the social network of group members to effectively target communication and contributions with greater efficacy. For example, Role identification, introduced above, describes the need to understand, adopt, and act within a defined role within a group. However, if one's identification with that role is not visible or otherwise obtainable by group members, any role that requires within group communication can not be fulfilled. If one self-identifies as a program manager but the rest of the group does not know it, that individual's ability to program manage could be adversely affected. Similarly, if one opts to become a mentor for new users but does not communicate that interest in some way within the group space, newcomers will be unable to find that user or benefit from that interest. So, Social awareness refers to the ability to build knowledge about the members of the group, Cultural awareness refers to the ability to build knowledge about how to constructively interact with those members, and Structural awareness refers to the need for the design of the space the group is working within to facilitate these, as well as many other, tasks.

## **Coordination activity**

### **Completing work required to coordinate group activity, including tool creation, task creation, structural design, using tools to mediate group activity**

Distinct from "Group awareness" is the notion of "Group activity." Whereas "awareness" as a dependency implies the need to understand group structure and culture to more effectively interact within that group, "activity" as a dependency is meant to communicate the need for effective interactions within that system. These interactions can be seen as potential coordination mechanisms for awareness (for instance, increasing awareness of group membership could be done through a published members list), however the means behind that communication will also require group coordination (for instance, continuing from the above example, how a "member" is defined, the frequency the list is updated, and the means by which those updates take place are all potential points where coordination between group members can break down). Further, this dependency is intended to be distinct from project activity, or the work that occurs which is the focus of a given group, as group activity is the work to maintain the group itself (ie, editing and maintaining the WikiProject pages themselves as opposed to completing the work which is the stated focus of the project). This distinction is important, as in CBPP communities task selection is typically done in a bottom-up, emergent fashion rather than the top-down task assignment more generally found in traditional organizations, requiring a far more nuanced view of the coordination required to facilitate these types of interactions. In other words, while awareness is related to context, development, and individual knowledge, Coordination activity is related to coordination between multiple group members, multiple projects, multiple tools, or any combination of these (and other) actors interacting within a system.

## **Social coordination**

### **Social activity within group (distinct from group goals), agreement, disagreement, mentorship, finding collaborators, engaging collaboration**

Within the activity to maintain the distributed and emergent coordination of the group is the notion of the social interaction related to maintaining the group culture. Recalling that Group awareness:Cultural is a dependency defined to categorize issues of awareness, this dependency implies the actual group interactions which inform and facilitate that culture. That is, beyond the awareness of a culture required to act within its boundaries is the distributed and collective interactions of those group members which, in aggregate, coalesce to comprise that culture. More broadly, and similar to Goal activity:Social, this category is not limited to simply “conversation” with other group members. This means that agreements, disagreements, requests for collaboration (RFCs), opinions, and discussions related to a project (even if the actor is not a project member) would constitute a “social” interaction, and would therefore be placed under this category.

## **Production activity**

**Activity to further group goals (generic, distinct from in-group activity), content creation, adding content, evaluation of content**

Goal activity is related to the work that is outside the coordination of the group, or of individuals, tools, bots, etc, that interact within that group. This is related to the work that individuals complete to further individual goals. These goals, though frequently motivated by individual desire, often align with group goals as well, facilitating a mode of interaction whereby these individual desires match group goals to the extent that continued participation may be beneficial to both entities, despite initial motivations. For this top level category, this may include generic contributions to article content or reverting vandalism (as watching for vandalism would be Group awareness, and as stated above, regardless of one’s personal allegiance to any group).

## **Social production**

**Communication regarding external artifact (non-group related, although it can be on group/project pages), socialization with external community**

This category can be distinguished between Group activity:Social by the nature of the communication. Where Social coordination is related to the communication between group members to either further group or individual goals (including coordination, policy, or gauging consensus among the group itself), this category is related to the communication around an individual work artifact - in our case, typically an encyclopedia article. This may include conversations about article content, article notability, article moves, etc. Namely, this category does not indicate that the “social” activity is purely conversational - instead it indicates that communication is occurring that may potentially benefit the larger group. And again (as stated above twice), this “group” does not have to be one with which an individual self-identifies. Rather, the “group” in that context indicates the larger community - of editors, of Wikipedians, and of readers.

## APPENDIX C: INTERVIEW PROTOCOL

The purpose of this interview will be to elicit reflection and greater understanding of existing coordination practices within voluntary virtual teams. For the sake of this study, voluntary virtual teams can be defined as small groups of distributed individuals engaged in a common task, typically for non-monetary purposes. Further, coordination is stipulated as the types of activities editors engage in to maintain an awareness of others' activities within that team as well as to communicate their own activities to others. More specifically, we are focusing on the capacity for tool-mediated coordination to increase the transparency of online activities and to more effectively facilitate efforts of distributed online team members to interact with one another. This may include tools for communication (ie, a synchronous communication system that allows for more immediate responses from online team members), tools for task management (ie, a system that would both allow for new tasks to be added and ranked, as well as tracking the progress of ongoing tasks), and tools for member management (ie, a system that would facilitate tracking active project members along multiple dimensions, such as explicit membership, activity-based membership, or interest-based membership).

The basic interview format will include three parts: biographical information, current coordination practices, and desired improvements for tool-mediated coordination. As this is a semi-structured interview, it is expected that most questions in the latter two parts will act as a springboard for the participant to reflect both on how they perceive current coordination practices as well as types of improvements they may desire to improve on current practice. The questions presented here are a starting point for the interviewer to encourage further reflection and dialogue by the participant. Interviews may be conducted over voice-based media (ie, phone, Skype audio call) or via text-based media (ie, messenger, instant message), and in both cases the interviewer will be expected to explore responses to gain greater insight into the participant's perception of both current and desired coordination practices.

### Part 1: Biographical Information

- What is your Wikipedia user name?
- How many years have you been a Wikipedian?
- What WikiProjects are you active in? And for how long have you been active in each of them?
- What do you feel is your primary contribution to these projects (ie, article editing, coordinating editors, updating project documentation, organizing categories, etc)?
- Do you think you'll continue to be a member of this/these projects in the future? Do you feel that you're having a positive impact on project goals?
- Do you think that these projects/this project has been effective in achieving its goals?

### Part 2: Current Coordination Practices

#### *Overall level of collaboration*

- How often do you work with other editors within or outside your WikiProject(s)?
  - What types of things do you collaborate on with these other editors?

#### *Tracking and finding members/collaborators (Member system & Reputation system)*

- How do you find other project members that may be able to help you with work, or may be able to complete work that you want done?
  - Task lists, Project Talk page posts, personal messages?
- How do you determine who would be the best person to work with, or to complete a task?
  - Ie, personal history, edit/transaction history, User Page, etc
- How do you judge the quality of the interpersonal relationships you build within WikiProjects, such as editors you've had interactions with in other contexts?

#### *Communicating with members/collaborators (Communication system)*

- How do you communicate with other editors that you work with?
  - User Talk page, Project Talk page, Article Talk page, email, IM, etc?

#### *Identifying work items (Task system & Rank system)*

- How do you decide what to work on next?
- How do you let other project members know what you're working on, and when it's complete?
- Given that most projects have many competing demands that vary in levels of time required to complete and overall importance, how do members of projects you've participated in decide what to do next? Is there a codified rating system or are most decisions ad-hoc?

### **Part 3: Desired Improvements for Tool-Mediated Coordination**

- For each of the above systems, what are some things that you think work particularly well? What are some of the things that don't seem to work?
  - *Why* do things seem to work or not work?
- For each of the above systems, what type of improvements can you envision which might improve coordination and collaboration between group members?
  - Would it be overall project knowledge, visualizations? Or a codified ability to weight the importance of tasks? Or an easier way to dictate or explore who is working on what, even if it's multiple editors on a single task?
- At the project level, what types of overall activities may improve the health and functionality of the project? What type of problems do you think facilitated these solutions (ie, is this a global issue or project specific)?
- At the project level, what types of overall activities may increase the project's ability to attract, onboard, and retain new members?

### **Closing & Conclusion**

- Thank the editor for their participation and give them the contact information of researchers involved in the study in case they have any future questions.

## APPENDIX D: EXAMPLE INTERVIEW

The following is the transcript of an interview with one of the WikiProject members, providing as a concrete example of how the interview protocol was executed in practice.

Note: All personally identifiable information has been removed from the transcript below.

---

04\_0002 : R.

04\_0003 : Ok, are you still there?.

04\_0004 A: Yeah.

04\_0005 Q: Okay.

04\_0006 Q: I just wanted to, before we get started, verify that you got the consent form.

04\_0007 Q: And you had a chance to read that over.

04\_0008 A: I did, it's fine.

04\_0009 Q: Okay and do you mind if I record this so we can transcribe it later?.

04\_0010 A: No, that's fine.

04\_0012 Q: Alright, well, i guess get started then.

04\_0013 Q: The goal that we have just in general for this project is to, at this point kind of figure out how people within WikiProjects are currently doing their work, how they're finding stuff to do, finding people to work with, and just sort of the broad spectrum of work that goes into the meta work, that goes into doing actual project work.

04\_0014 Q: Then moving forward from there, trying to understand what are some of the details within that, and even beyond, what are some of the ways we can try and improve these current structures.

04\_0015 Q: So that's the general kind of gist of the line of inquiry we're on right now.

04\_0016 Q: And I guess just to get started, if I could ask you what projects you're currently working on.

04\_0017 A: So i've been in a few and I am currently a member of a few, all of them are on the English Wikipedia.

04\_0018 A: I'm a member of the Piracy project, WikiProject Occult and WikiProject Skateboarding.

04\_0019 A: So, quite a diverse range.

04\_0020 Q: Definitely.

04\_0021 Q: Let's see, so for each of these kind of project, is it more, what sort of is your driving force for trying to contribute to them.

04\_0022 Q: Is it just personal interest? Why did you choose this kind of mix?.

04\_0023 A: Yeah, so, perhaps to start with a bit of an explanation.

04\_0027 A: But I suppose my natural interests would be in politics and I suppose cricket and football.

04\_0028 A: Cricket might be a bit alien to you guys, but I try and avoid working in those areas.

04\_0029 A: I kind of don't edit stuff related to cricket at all.

04\_0030 A: I avoid editing things related to football, and I try and steer clear of politics, just because I never know where I'm going to have a conflict of interest or if I'm going to have some sort of personal view that might cloud my contribution to otherwise true context.

04\_0031 A: So I've tended to gravitate toward projects that have personal interest to me.

04\_0032 A: I do a lot of stuff in history and religion, though I'm not a member of the history WikiProject, except in peripherally and likewise with the Catholic WikiProject, which is mostly inactive anyway.

04\_0033 A: So the skateboarding is something I did when I was young, and so as a something that had a personal interest to me, I kind of joined it and put a bit of a fitting there.

04\_0034 A: The occult stuff is again something that interested me and so I'm into that.

04\_0035 A: And likewise the piracy was personal interest to me, sort of doing a lot of stuff in the 17th century history area and so fell into that as well.

04\_0036 A: So more than anything, personal interest.

04\_0037 Q: So kind of personal interest that manages to not overlap with your professional responsibilities or life.

04\_0038 A: As much as possible yeah.

04\_0039 A: I mean, yeah as much as possible I try and sort of avoid those area, I have sort of actively avoided joining projects that would encourage me to edit in those areas as a result.

04\_0040 Q: So you had mentioned that you did some editing around, I think it was history, but you're not currently a member of that.

04\_0041 Q: Do you think that your membership in say the piracy, occult and skateboarding, just the membership alone in those projects changes the nature of how you might contribute to articles within the scope of those projects?.

04\_0042 A: Yes but not sort of in the sense of neutrality or anything like that.

04\_0043 A: Especially in wikiproject piracy, it's either historical fact or it's not, so there isn't really any nationalist conflict of interest or one group of pirates over another, so it's kind of an area where there aren't a lot of conflicts in terms of context.

04\_0044 A: A lot of it is textbook history stuff.

04\_0045 A: I suppose the reason I'm not a member of the history WikiProject is that piracy is kind of a, it's not directly I think one of the subprojects, It's kind of a child project of both Military History project and the History project, and the Military History project is itself a child I think of the History project, so it didn't seem necessary to join all of them.

04\_0046 A: And a lot of the history stuff is actually, I tend to stay more focused on kind of ancient history or modern history and and it's the whole bit in the in the middle, so there was less of a reason for me to jump in there.

04\_0047 Q: Okay, well I guess so I was just kind of curious if there was something distinct about the structure of a Wikiproject that would mean that your contributions to say WikiProject piracy is going to be different from a project you're not necessarily a member in, or is membership not necessarily modifying the types of interactions you'd have within those areas.

04\_0048 A: I suppose it has a little bit, I mean the WikiProject skateboarding is probably, I'm probably more active in trying to invite people into that WikiProject, though not necessarily successfully.

04\_0049 A: There's a lot of new editors that kind of show up, given the nature of skateboarding as a cultural thing.

04\_0050 A: There are lots of young, new editors or accounts of editors that kind of show up and edit one page and then they disappear and I think given the potential for some of those WikiProjects, there is a potential for them to become more involved if they're invited to the right places and supported by the right people, so in that sense, my interaction with people in the context of WikiProject skateboarding is probably more tolerant of editors who show up and make mistakes or what have you to articles related to skateboarding because my first thought is often that I hope that they can become an active editor as opposed to this is just a young school kid who just wants to vandalize something.

04\_0051 A: So in that sense I'm more tolerant.

04\_0052 Q: That is kind of interesting, that brings up another points, which is generally how you perceive your role in some of these projects.

04\_0053 Q: You mentioned in skateboarding you send out invitations, so to what extent would you say your role is kind of like some of the organizational work or the project work versus article editing work or any other kind of types of works that occur within those projects.

04\_0054 A: I suppose as I've seen from some of the larger projects, projects that I'm involved in are very small.

04\_0055 A: There are maybe half a dozen to a dozen active members to each of the members I'm involved in.

04\_0056 A: It's not like the military history project where every year they hold elections and coordinate roles and have meetups and interactions with live reason.

04\_0057 A: Ours is probably a lot more informal, I suppose I've kind of naturally kind of taken up a leadership role, sort of in that way.

04\_0058 A: Bear in mind we're only talking about a few people, but people have sort of come to me on occasion sort of looking for guidance or advice with regard to either technical things or related things, but none of those projects is very sort of formalized, leadership or coordination structure.

04\_0059 A: It's all very informal.

04\_0060 Q: Okay, and then as far as in general and subjectively of course, do you think the projects you're involved with are you know, mostly successful?.

04\_0061 A: Yeah I suppose it's relative isn't it.

04\_0062 A: Were they bigger with more people with more interest, they would be more successful, but they did what they're supposed to do, which is to bring a handful of people together and to some extent coordinate some of the work in those topic area.

04\_0063 A: So yeah I mean if you judge it in terms of popularity or number of people involved or number of edits, or just topic area articles, we're not particularly successful, none of them are.

04\_0064 A: But if the goal is just to sort of provide a platform for a few people to occasionally get together and talk about a particular issue in a subject area, they've all done that successfully.

04\_0065 Q: And I think success is definitely not membership counts, a lot of it can be member tenure and member involvement and the degree you can have some sort of positive effect on articles.

04\_0066 Q: So there's a lot of different areas there.

04\_0067 Q: But very frequently we do hear about the most successful projects are the biggest ones, and to a certain extent that could be true I suppose, but it's not the only way.

04\_0068 A: Well I think some of those larger projects have the capacity to just as a matter of sheer numbers to as a coordinated effort start a series of articles in a particular area, you know at the same time, and collectively work together to source them or cite them or collectively work together to get them all to featured article status or at least good article status, so as a collective unit they might have all done something small, but together they've managed to produce half a dozen new articles in whatever time frame they've set for themselves.

04\_0069 A: I've seen that done quite successfully, but the projects that I'm in haven't been close to those sorts of things.

04\_0070 A: If I want to write an article or if I want a project to write an article, it's usually then that I go and write the article, and that's fine.

04\_0071 A: There's a couple of other guys in each of those projects who are content producers and will go out and do the same thing.

04\_0072 A: But it's not on the same scale.

04\_0073 Q: So when you are trying to put together an article or something, how do you go about doing that, trying to engage other people? What's the specific mechanism that you use to say, "Hey, come over here and let's work on this together.

04\_0074 Q: [14:58].

04\_0075 A: Yeah, or sometimes I'll post on the talk page of the project, or contact individual members in particular and say, look I know you're interested in this [inaudible] did you want to contribute to this or I've started this article, have you got some thoughts? And on occasion if I start on article in an area that's not related to a project I'm in, I'll kind of do the same with other projects, so if I say, I don't know, food or architecture, I'll go to the projects and sort of leave a note on the project food talk page or something like that, just to say look, I've started this article if anyone's interested.

04\_0076 A: And I do the same thing.

04\_0077 Q: So it's mostly these sort of direct engagements.

04\_0078 Q: How do you identify people? You said it was prior editing history of the folks that you talked to.

04\_0079 A: Yeah, so on most projects or all the projects that I'm a part of have a membership list, an established membership list on the front page if you go through those projects, where members are registered, where members are able to register their interest in your project or register as members of the project.

04\_0080 A: If obviously if they're on a watchlist or I see somebody put themselves forward as a member of a project, I'll go and introduce myself on their talk page and say hello.

04\_0081 A: Or if they, if I can see someone editing a lot of pages that are on my watch list or on the watch list of the project, I'll often go and say hello and either invite them to join [inaudible].

04\_0082 A: just have a talk with them about what they've been doing so, the role of people, whatever might be going on.

04\_0083 Q: Okay, so let's see.

04\_0084 Q: Would the, one of the other things that I was.

04\_0085 A: Uh sorry, I don't know when that was, but Greg posted a question about the role.

04\_0086 A: Did you wanna.

04\_0087 A: ?.

04\_0088 Q: Oh yeah, sorry.

04\_0089 Q: Let's see here, so within these projects, as you said you've taken a sort of leadership role, how would you say the tasks are doled out to the 6-12 members? Excellent question Greg!.

04\_0090 A: Yeah, of the three options, it's definitely the latter, it's not a very formal process, if, often, so I'll take WikiProject skateboarding there are a few active participants.

04\_0092 A: And some of them have, en, you know because I'm in history, and some of those is on states, a couple of them are about Europe, and there is a couple of across the US, so naturally if it's something that's common US based or related to the history of a particular US company, then, OK, one of the US guys to have a look at the article work to highlighting that.

04\_0093 A: En, took within, en, often most pages are on the watchlist anyway, and so we will are one the article talk page or on the project talk page or personal talk page.

04\_0094 A: So it's not really that, you know, list of things to do for whole time, it's just kind of naturally fall into those UK business related to United States and some of US business might get better understanding than the others I do.

04\_0095 A: whereas we really simply are drive, if you like, creating articles, a bunch of stray, and skateboarders.

04\_0096 A: and some of them will create, some of them not.

04\_0097 A: and yeah, that was mostly me recreating those articles and then adding stuff but yeah, it was underneath the formal guide, these are the ten things you have to do, you have three, you have five, ahaaa.

04\_0098 Q: Yeah, yeah, so when you guys sort of hashing out the work that potentially could be done, is that's em, so Greg I'm sorry, just saw another post behind, en, so when it comes to sort of hashing out that process of different texts that are available, what areas do you guys actually use to make those decisions, not just that structure thing, or necessarily even like a democratic process, but do you have conversations, or is it really just a go-it-alone kind of process, is it email, and en ,what kind of, you know.

04\_0099 A: Well I can tell you it's almost 100 percent talk page based, we will tend not to email each other, probably, we don't sort of use IM or chat online those kinds of, because we are sort of around the world, anyway, but the discussions we have normally are entirely on the talk page, we could say that they are sort of one-the-talk-page project, so on each case, probably, all of that are you know, what tends to happen is probably a far moral getting precedence, somebody would have an article they've created or an article that they are sensitive so it's on their watchlist, and a new article we created by someone in the group that only act to the article watchlist, and radically could be created by someone, article, watchlist, so he would noticed that an article has been created, and as I said, there is being part of project most successfully chance of actually content creation, that's probably a couple years ago, there was really lack of clarity there were whole bunch of articles that were just basically drive by opinions and shout out articles and never come back, they went unsourced, they went and hands all right, and just kind of conversationally, and all of what is happening sort of drag back to the manus of star that sometimes keep me spanning, and there is a lot more sources, a lot, so it's, cause there's just so naturally way of variable and that still leads, still succeed to grow, so between the few people the project it is not really being a part of a least from my perspective, two parties, say, of common insight, OK, let's chat, just seems to do, basically any page, any article that has been tagged, has been related to this subject area, could do this work , so it's kind of by ... to work on something.

04\_0100 Q: OK, yeah, so and and a lot of ways it's just more of opportunity stake, en, you know, the task matches the time.

04\_0101 A: Very much, yeah, and I don't know that's lot, very different, and then, you know, some of the projects, have set of some articles, recently wrote completed by the most of them are you know and they have some of been drive to get those articles from stop to you know they been be especially a half of the articles are full of sort of.

04\_0102 A: aren't which is probably part of the reason why we have to actively recruit to the project because it starts to many many they have.

04\_0103 Q: That actually bring up another question, which is how do you, en, you mention that you can, you know, see that user has edited on biography living person scapot article, and maybe you send out an invite, how do you try engage people once they are a part of the project, en, because there is always the problem where someone will make like, they will join the project, add their names in the member list, they made one edit, and then they will never come back.

04\_0104 Q: Are there any specific strategies that you can think of that you currently do now within these projects, that basically trying to get people stick around?.

04\_0105 A: No, no really, I mean, as I said before, a lof of people that will join, they are kind of one-edit editors, who now are readers, and they see if probably with a particular article, so they can fix that particular article, and never engage them, they are on their talk page, a welcome note, and they only write on the project.

04\_0106 A: Sometimes they will join, and sometimes they won't, actually they will become acting, more than acting as members of the wikiproject, that kind of secondary, so in a lot of cases, we actually talk about acting pertinence, which is this is who before, sometime they lost, hahaha, and that's fine, you know, we know it's the case, especially in that specific subject area, but you know, I think, partly of them is trying to encourage them to stick around, not just that subject, maybe that subject area.

04\_0107 Q: Yeah, I mean, part of it is just en, you know, mentorship or training to become wikipedian, and be able to contribute to that in a productive way.

04\_0108 Q: Let's see here, I think just kind of moving on from that, the kind of final portion is just talking about some of the different types of improvements you could see, and this is more than just a daydream, blue sky, type of thing, like what sort of changes might you do in order to improve both the projects that you work on, and then just wikiprojects in general, for instance, when it comes to be able to track membership, which, you mentioned, is

sometime a little bit problematic because people, they are typically out of date, a lot of times, and don't necessarily accurately depict a project's membership.

04\_0109 Q: So those types of things, for instance.

04\_0110 A: Yeah, well, I supposed I haven't really been a member of a project, any one project.

04\_0111 A: It kind of like, in the case of [wikiproject skateboard](#), I kind of know the active members, because the inactive members aren't just the inactive members of the project, their inactive appearance, at least they don't edit more, so they've retired or have just moved on to something else.

04\_0112 A: In the case of some of the others I don't necessarily know all the people who will participate, but I've a pretty good relationship with some of the members, some of the who are active in an area, but are not members.

04\_0113 A: But I mean tracking membership when you talk about a group of active members, that you could probably count online, it's not as as other things I suppose.

04\_0114 A: We are not gonna probably talk about each project, I would could probably you know it's not problem, it probably talk about tracking them it that was probably more than half year, we could you watch is maybe considered for selective, and I mean, I would probably get little so you know, things happen really tow but as I say if you refused to say it's very different.

04\_0115 Q: And I guess kind of related question would be do you use any tools that aren't necessarily part of the core media wikiproject form, in order to do your project work? And what kind of thing would that be?.

04\_0116 A: Not really, And I suppose in the context, I'm not 100 percent what you mean.

04\_0117 Q: Like for instance, there is [Twinkle](#) or [Huggle](#) or auto wiki browser some of the user scripts statics that are there, en, some of the automated editing tools or semi-automated editing tools.

04\_0118 A: Em, I've , the [twinkle](#) I used , but it's not necessarily just for the project, it's just a common tool that I use.

04\_0119 A: I suppose that the context of the projects, so, these are few automated things that I set up, the [skype](#) try to react or there are any sort of automated tools, just cut off the area of the page, the talk page, that's pretty much it, and edits templates, that, you know, the classification fringe of subjective articles, that's pretty much it.

04\_0120 A: But we have edited, so we now have articles isolation tools, one of the tools, I couldn't tell you which one, but it basically sort of alert you to look at an article alert thing, the whole theory of the world, but anything that's in the subject area, so if something is not edited manipulation, or something is not edited under permission, to a higher status, all of those things, that's kind of pages that could be watchlist that track those things.

04\_0121 A: So everytime the tool edits that article, that particular page, we get a notification, you know, the page where wikiproject says working this change, articles that have been manipulation.

04\_0122 A: So you know, that's kind of tool that allows us to be a little bit more organized, if somebody has to do it manually.

04\_0123 A: That's generally why we have to do those things manually.

04\_0124 A: But there's, you know, we use those sort of tools, just to track things, but en, nothing that's kind of project specific.

04\_0125 Q: OK, so sounds like for the most part, a lot of the correlation and communication and everything that among the project is just purely done on the project talk page.

04\_0126 A: En, yeah, we are all sort of multiple people.

04\_0127 A: A lot of the talks are just conversations between two editors about particular subject will be on that project talk page or just the personal talk page, en, and on occasionally, the conversation has been related to the subject area, but has sort of correlated on the personal talk page, en, to something that ratify the project, we will then copy and paste that discussion over to the wikiproject talk page.

04\_0128 Q: Let's see, I think just kind of one more questions for me, and then I will see if the other guys here have any thoughts, would be just if you could talk to any of the but designers, or tool designers, or any of those guys who are working on the platform and to request something just to make your life easier, what would you ask?.

04\_0129 A: In the project content?.

04\_0130 Q: In the project content, yeah.

04\_0131 A: Ah, I mean, I supposed that what has been discussed before, and more in [quantex](#) to wikiproject parse, there are recognizing key words, there are inappropriate peaks, maybe thirty keywords, what would be great is that there is a tool that allow talk pages of new articles to be automatically.

04\_0132 A: [Stuart](#) [36:00 - 52:51].

04\_0135 A: tagged in the [WikiProject](#).

04\_0136 A: Not the article, the talk page.

04\_0137 A: Where the article contains [garbled] words that are, key words that are [garbled] but that's not necessarily something you could do, because of the project schedule.

04\_0138 A: You know, the things that would be relevant, like "park," or "skate," well, even "skate," that isn't specific to skateboarding.

04\_0139 Q: Mmhm.

04\_0140 A: That wouldn't necessarily work with that project, but for Piracy there are probably some really simple keywords, and the combination of any two of those.

04\_0141 Q: Mmhm.

04\_0142 A: ...correctly, but you would probably have an accuracy rate of 90-[garbled] percent.

04\_0143 A: [Laughs].

04\_0144 Q: Yeah.

04\_0145 A: And it would just allow the project to be instantly aware of new articles in an emerging (?) space.

04\_0146 A: So if somebody [garbled] where the project exists, an article about something that's related to the subject, especially given that WikiProject Piracy has such an extensive, I think there's about a hundred, I think it's almost two hundred, and they all sort of requested articles to create.

04\_0147 Q: Mmhm.

04\_0148 A: So, anyone obviously creates an article on that list, [garbled] or if it's something that's necessary to have consensus from people to automatically tag things,.

04\_0149 Q: Well, I'm always curious about how the presentation of information can impact how a project has actually worked, or how a project works, and the type of tasks that that project can do.

04\_0150 Q: So with something like that, I can totally see if you even got close to 90% of accuracy on that, then it's just more of a human process.

04\_0151 Q: You can go ahead and create different tasks for project members to go through, and tick them off one at a time, even with six or a dozen people.

04\_0152 Q: Something that seems so unmanageable at first can become more manageable.

04\_0153 A: Yeah, all of the tools that I could possibly suggest would all come from a place of very small projects.

04\_0154 A: [Laughs] Not that I have an objection to the large projects, but I just haven't got to be a member of them, so I know what does and doesn't work for small projects.

04\_0155 Q: Mmhm.

04\_0156 A: And the tools that I might suggest for one of my projects might be completely superfluous to larger projects.

04\_0157 Q: Mmhm, yeah.

04\_0158 Q: It's a different context.

04\_0159 A: One thing that I've thought about previously.

04\_0160 Q: Okay.

04\_0161 Q: Well, I think that's most of what I had to ask.

04\_0162 Q: I'll see.

04\_0163 Q: Greg and Ruiyi, do you guys have any questions?.

04\_0164 A: There's one more, which is kind of cross-wiki, I don't know if you're focusing on the ... things, or relations with commons(?), but there seems to be quite a large set of media on Commons that--a lot of those Commons projects are either articles on the [garbled], so there's just vast amounts of content that are on Commons and nowhere else, and the creative commons license that gets used on Wikipedia, and, but obviously the naming conventions there are different, the meta-tagging (?) is different.

04\_0165 A: You end up with a lot of stuff that--obviously, for WikiProject Piracy, there's a lot of stuff that's related to piracy in the Caribbean, in that region, that is, French, Dutch, or Spanish--the English were there, but it was more proportioned, and so a lot of the media, the images and paintings and pictures, and to a lesser extent photographs, a lot of the tags are in Spanish, or French, or German, or Dutch, and it's really quite difficult to guess what media might be relevant to an article you're writing.

04\_0166 A: So if I'm writing about a particular conflict, I really don't know [garbled] pictorial representation somewhere on Commons.

04\_0167 A: And vice-versa, sometimes I go to Commons, I go through some of the media there that doesn't have articles, things related to the subject areas that I'm interested in, I think that I should create an article just to use some of these.

04\_0168 A: I don't know how a tool would work in that regard, but it strikes me as being--it would be advantageous if something was available to project members to try to identify available [garbled] on Commons.

04\_0169 A: There is a “files” [garbled] English Wikipedia, but as far as I understand it’s only for files that have been uploaded to Wikipedia that don’t have articles there.

04\_0170 Q: Yeah, I’m not sure exactly how that might work, but I have seen some scripts that look at inter-wiki links between Commons and foreign language wikis, so say the German or the Russian, and it does seem like if there was a corresponding article in, say, the German Wikipedia and English, that if there was something on Commons that was linked to one and not the other, that would be one way of trying to put that together.

04\_0171 Q: But that definitely sounds like it could be a pretty sticky problem.

04\_0172 A: Yeah, it’s not a problem that’s going to break things, it’s just sort of--if there was a way of getting more benefit out of that, that would be great, but there isn’t.

04\_0173 A: [Laughs] We’re kind of in that position now, we’re not going to lose anything from it, but--.

04\_0174 Q: Of course, yeah.

04\_0175 Q: It’s just an interesting starting point, where if you can put information in front of a project in a way that’s useful to them, no matter what that information is.

04\_0176 Q: This is why it’s always kind of a curious thing to think about--daydream, essentially--about the types of things that might make your life a little bit easier.

04\_0177 A: Yeah, and there’s a secondary part to that.

04\_0178 A: But I don’t know if other languages of Wikipedia have WikiProjects, but it strikes me as self-explanatory if a topic was of more interest or relevance to--if Piracy was a project on, say, the Spanish language Wikipedia, it would almost by default be of interest to them.

04\_0179 A: Which language(?).

04\_0180 A: Not speaking Spanish, I can’t go over there to the Spanish Wikipedia and find out if they have a project, but I don’t know if it’s something that even happens, whether there’s some collaboration between WikiProjects, or if other languages have WikiProjects.

04\_0181 Q: I believe they do, but I think collaboration is difficult to a certain extent going from other languages to English, just because English language Wikipedia is so much larger.

04\_0182 A: Yeah, the articles that are missing would be kind of [garbled].

04\_0183 A: That said, I know that there are plenty of articles on other language Wikipedias that we don’t have, so.

04\_0184 A: [Laughs].

04\_0185 Q: Well, it’s definitely interesting, because that’s another one of the possibilities that could be fruitful.

04\_0186 Q: It could be something where the information is already there, we just have to dig for it properly.

04\_0187 A: That’s definitely right.

04\_0188 Q: This is actually great.

04\_0189 Q: I think, again, I’ve got pretty much everything, and it sounds like Greg has got everything, so unless you have any other comments or thoughts, I’d just like to thank you once again.

04\_0190 A: Oh no, it’s fine, and let me know if you need any follow-up or anything, if you need to revisit any of those things.

04\_0191 A: It’s fun to sort of think about some of those things.

04\_0192 Q: Yeah, great.

04\_0193 Q: Well--.

04\_0194 Q: Ruiyi (Ru): Chinese Wikipedia doesn’t have WikiProjects.

04\_0195 Q: The Chinese Wikipedia doesn’t?.

04\_0196 Q: Ru: Yeah, it doesn’t.

04\_0197 A: Is that just sort of a function of, they don’t have them because there are restrictions on online collaboration, things like that? Is there a particular reason that there are no WikiProjects, or is it just kind of.

04\_0198 A: they don’t?.

04\_0199 A: Ru: I don’t know the reason.

04\_0200 A: But the Chinese version of Wikipedia, it does work, people collaborate in Chinese, but there is no project on Chinese Wikipedia.

04\_0201 Q: Hmm.

04\_0202 A: That’s interesting.

04\_0203 Q: We’re going to have to take a closer look!.

04\_0204 A: [Laughs] I’ve created more workload!.

04\_0205 Q: [Laughs] Exactly.

04\_0206 Q: I might try to reach out to you at some point, because we are actually trying to put together some things that--hopefully--can make some lives easier.

04\_0207 Q: But for now it’s mostly just trying to understand and make sure that we’re doing it in the proper way.

04\_0208 A: That's good.

04\_0209 A: I'd have to find them for you, I don't really remember the leadup to that, but there was a place, I think it's on stage at the village pub, for a set of user rights--so projects can see their user rights.

04\_0210 A: [Garbled] It wasn't particularly popular, it was way too hard to implement, but the idea was essentially a set of tools for active project members, and they could be given the tools by agreement of all the other project members [garbled] But in that proposal there was, I think, tools proposed on the basis that they would be useful to WikiProject members in general, not just WikiProject [garbled].

04\_0211 A: So I might be able to [garbled] if I can find that for you.

04\_0212 Q: Yeah, that could be interesting, I haven't heard of that.

04\_0213 A: It was never going to get consensus, because you were effectively making project-specific [garbled], you could sort of protect pages with their own project areas, delete stuff, [garbled] under the banner of a project.

04\_0214 A: [Mumbling] [Laughs] The actual discussion, as I recall, contained some interesting [garbled] that people wanted, that said, for different reasons, so if I can find that for you, I'll send a link.

04\_0215 Q: That'd be great.

04\_0216 Q: I'll take a look too, because that actually sounds pretty intriguing.

04\_0217 A: Yeah, I remember the proposal itself was overwhelmingly shut (?) down, it was kind of a chance for people to kick some ideas around.

04\_0218 A: I don't know that they're actually -.

04\_0219 Q: Mmhm.

04\_0220 Q: And it's interesting to see what project members at the time thought might be important, or, you know, dangerous to the community.

04\_0221 A: I mean, I think it was overwhelmingly considered inappropriate, sort of project specific editing or something along those lines (?), for obvious reasons.

04\_0222 A: [Laughs] But some of the other things, the topics of conversation--.

04\_0223 Q: Yeah.

04\_0224 Q: If you do find it, that would be great.

04\_0225 Q: If you could send it my way, I'll take a look.

04\_0226 A: Cool.

04\_0227 Q: Well, alright! I guess I'm gonna wrap up here.

04\_0228 Q: I think we're all set on our end.

04\_0229 Q: Thanks so much, and I'll keep you posted if there's anything else.

04\_0230 A: No problem, any time.

04\_0231 Q: Alright, have a good night.

04\_0232 A: Thanks very much, see ya.

04\_0233 A: All: Bye.

## APPENDIX E: TESTING CODER AGREEMENT OF OUR THEORY OF COORDINATION FOR COMMONS-BASED PEER PRODUCTION COMMUNITIES WITH SCENARIOS

During theory evaluation, iterations of the codebook for our theory of coordination were evaluated and continually refined through multiple codings of scenarios drawn from interview transcripts. Through each iteration, disagreements between coders were resolved through adjudication and the codebook was updated to more accurately reflect varying interpretations of how interactions with the community aligned with current understanding of the theory of coordination framework.

Adjudicated Dependency	Initial Code	Action (Creating tools to support project work)
Coordination activity	Creating project pages	000 - Initial creation of project space
Group identification	Joining project	001 - Self-identify as project member
Social coordination	Setting up project	002 - Working with members to create project guidelines
Coordination activity	Facilitating project interactions with tools	003 - Creating tools that facilitate enforcing project guidelines
Social coordination	Validating tools	004 - Work with members to validate tool functions
Group identification	Working towards project goals	005 - Manually invite newcomers to the project to complete project goals
Social coordination	Helping newcomers	006 - Answer newcomer questions about how to edit articles
Social coordination	Planning future tools	007 - Work with project members to plan new tools to better facilitate project activity
Coordination activity	Creating tools to facilitate project interactions	008 - Independently create new tool to invite newcomers to the project
Cultural awareness	Learning how to support argument for existing tools	009 - Monitor tool discussions to determine the best way to interact with admins
Coordination activity	Creating legitimate bot account	010 - Post to bot request page to have existing tools be made legitimate
Social coordination	Communicating need for tool	011 - Support argument with community administrators to justify tool usage
Social awareness	Watching for project work	012 - Monitor project pages for new questions from newcomers
Social coordination	Answering article question	013 - Answer newcomer's question about editing a specific article
Role identification	Finding mentee to collaborate with	014 - Identify as mentor for newcomers
Social coordination	Working with mentee	015 - Work with mentee to teach editing best practices
Social production	Editing an article with mentee	016 - Work with mentee to enact those best practices editing a specific article
Social production	Working with mentee to edit article	017 - Answer mentee questions about editing an article on the article talk page
Group awareness	Monitoring mentee edits	018 - Add the article to watch list to monitor potential future edits by mentee

## APPENDIX F: INSTRUCTIONS FOR TASK DECOMPOSITION AND DEPENDENCY MAPPING TO CREATE CONCEPT MAPS FROM SCENARIOS DERIVED FROM INTERVIEW TRANSCRIPTS

The following illustrates the instructions that were given to the research group before coding the WikiProject scenarios (for example, see Appendix E for a snippet from the tool maintenance scenario).

---

To more effectively map real processes to the potential dependencies which could impede the success of those processes, we will be using examples from interview transcripts to decompose tasks required to reach some common goals within WikiProjects. Through this practice, we hope to illuminate the complexity of common tasks, highlighting the need for continued development, awareness, and collaboration to most effectively contribute to and participate within commons-based peer production systems.

The activities that we will be examining have been taken directly from the stories we have been told. They represent real activities that will act as the seed for task decomposition - the deconstruction of the stated activity or goal into its constituent parts, including the context that led up to the task, the development required to complete that task successfully, the process of attempting to complete the task itself, and the means by which that task can be re-integrated back into the whole. For the sake of uniformity, I will complete the task decomposition for each task. Dependencies that can be identified within the decomposed task can then be utilized as a means of interpreting the trajectory of common activities, including pointing towards problems that may arise and possible means of resolving them.

What follows are a few of the common tasks we have identified through the course of interviewing WikiProject members, a brief description of each task, and snippets of the transcripts in which the task was introduced.

### 1. Generating audio for WikiProject Spoken Wikipedia:

The Spoken Wikipedia WikiProject ([https://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Spoken\\_Wikipedia](https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Spoken_Wikipedia)) provides a location for editors to collaborate on creating audio recordings of existing Wikipedia articles. These articles can then be accessed by those who understand the spoken English language but not the written language, those with visual impairments, or those who just want to learn about a subject without sitting down in front of the computer. Some of the challenges faced in this project include the need to identify articles that require a spoken version; preparing the equipment to record that article, ensuring that pronunciation is accurate and quality is acceptable; and finally (among other challenges), ensuring that the audio version of the article is up-to-date. After all, once the audio recording is made it can quickly become out of sync with the text version, especially in the case of an active article or more topical subject matter.

*08\_0287 A: I haven't tried to get anyone to join a WikiProject, I mean, I guess the closest thing would be.*

08\_0288 A: *there was some London event where they had a museum, they had some kind of Wikipedia Museum Night or something, **and one of the things they wanted to set up there was a recording booth so people could record articles**, so they came to WikiProject Spoken Word for advice about how to do that.*

08\_0289 A: *My thoughts on the matter were that it was kind of impractical, because **recording an article is incredibly time-consuming and somewhat difficult to do.***

08\_0290 A: ***I recommended that they have people do pronunciations**, because those should be rapid-fire, just get them right in there.*

08\_0293 A: *But you know, **myself and one or two other people tried to work with the coordinators**, and I think we did to some extent work with the coordinators, to help them work out some of the technical details, what they were getting into.*

08\_0294 A: *And **I wouldn't be opposed to doing something like that, if I were in New York or San Francisco**, you know and there was gonna be a critical mass of Wikipedians so that I could organize something like that, or be a part of it.*

08\_0296 A: *That, that would be appealing.*

## **2. Dealing with copyright infringers:**

The Copyright Cleanup WikiProject ([https://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Copyright\\_Cleanup](https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Copyright_Cleanup)) provides project members with a platform to identify and combat potential copyright violations within the encyclopedia. As an “encyclopedia that anyone can edit,” from those driven by altruism, curiosity, or a desire to share personal knowledge, to those who aim to contribute with less noble goals, potential copyright violators are both often difficult to detect as well as hard to deter. Beyond identifying copyright violators for purposes of tracking or deterrence, identifying copyright violations among the millions of articles on the English Wikipedia alone is a complicated task, requiring constant evaluation and monitoring of a broad spectrum of articles. The complexity of this task, and the dearth of tools available to mediate some of these problems, indicates a process that is rife for observation and intervention, as the following transcript snippet shows.

07\_0398 Q: *But when you talked about how you would go about identifying the different types of work that you do [identify and remove potential copyright violations], it was generally the lists that are on the project pages or requests for missing articles - Does that mean that you don't use those watchlists as extensively? Or that you just have other different ways of going about that type of work?*

...

07\_0401 A: *But, **when I create an article after copyright problems**, I usually add it to my watchlist for just a short period of time, and then I will take it off my watchlist and let it develop independently.*

07\_0402 A: **So I will use watchlists for copyright problems that are recurrent, or that are of the sort that have red flags that tell me that they could come back** - for example, school articles, a lot of students like to add information to school articles, and many of them don't have the skills yet to write original content so they'll just copy/paste from official publications.

07\_0403 A: Or television episodes, oh I hate those.

07\_0404 A: People will copy plots from TV guide.

...

07\_0412 Q: But, kind of related to that, is there any other type of thing that you think would make this easier for you?.

07\_0413 A: Well, **I could imagine all kinds of tools that would help**, tools that would make it easier to figure out if a person who has received copyright problems, notices, like **one of the things that you should do when you find an article that has copyright problems is check the history of the contributor to see if they've been notified before**.

07\_0415 A: That CC guide list that I showed you, the one where we have a **list of unfulfilled evaluations dating back to 2010, if we could have caught some of those people earlier then there might not be thousands of articles created by them that need be evaluated for copyright problems**.

07\_0417 A: So, you know, **a tool that would tell us this person has been flagged for copyright problems twice before would make it easier for us to figure out how to work with them**.

07\_0418 A: Um, you know, so, I can imagine all kinds of different tools that would help this work function better.

07\_0419 A: From streamlining notices to processing, to especially assisting with investigations.

### **3. Finding collaborators:**

While many group members select tasks based on a whim and complete them largely in isolation, it also occurs that group members need to seek out assistance from others to complete their work. This can take many forms, from asking questions on the project talk page, the article talk page, or seeking out specific users based on their prior activity or existing knowledge. For this scenario, we'll take as an example the snippet from the transcript given below, in which the editor discusses their strategies for identifying and engaging potential collaborators in editing work.

05\_0111 Q: Kind of similar, do you normally do article editing work yourself? You mentioned that you sometimes asked questions on [WikiProject] - is that normally to find collaborators or get a one off answer, or something else?.

05\_0112 A: For me, **I usually use [WikiProject] to get a specific answer**.

05\_0113 Q: *Since part of what I'm focusing on is online teaming, I'm curious to what extent WikiProjects are collections of individual editors versus more group-driven collaborative processes.*

05\_0114 A: ***There is one editor I feel very comfortable asking questions of and we usually communicate via email now.***

05\_0115 A: *This person also has a PhD in plant systematics, so it's very good working relationship.*

05\_0116 Q: *So it's again something that kind of comes with experience, knowing the movers and shakers within the project?.*

05\_0117 Q: *Was that experience gained mostly through co-editing in articles, visibility in project, email list, or ??.*

05\_0118 A: *I think a lot of people do what I do, **edit a topic they like and then form their own loose groups to work on articles**, using people they get along well with.*

05\_0119 A: *yes. [in response to 05\_0116 above]*

05\_0120 Q: *It seems like a semi-structured process, just curious about the actual human interactions that lead to the semi-structure.*

05\_0121 A: ***for me co editing.*** [in response to 05\_0117 above]

05\_0122 A: ***I'd stumble across someone on an article I was working,*** talk to them on the article talk page or their own, and then when comfortable with them, I'll email them, if they have email turned on.

05\_0123 Q: *Ok, that makes sense.*

05\_0124 A: ***wiki is very unstructured*** and yet if you get caught up in a big problem, like getting blocked, hauled to arbcom, etc, it's very stiff and bureaucratic.

#### **4. Creating tools to facilitate project activity:**

One way that many projects are easing the burden on individual project members is by adopting tools that facilitate project work. This could include setting up article alerts that notify project members when an article changes status (for instance, receiving a nomination for Featured Article, or being nominated for speedy deletion), or potentially to create dynamic member lists based on the level of activity within the project space. The specific location for this scenario is within the Teahouse, the Q&A project intended to help newcomers become integrated into the Wikipedia fold in a friendly and welcoming manner. One distinguishing feature of this project is that it might be considered a task oriented group rather than a topic oriented group. For instance, while members of WikiProject Birds may consider the primary goal of their project work to be to improve the quality of articles related to birds, members of the Teahouse consider the primary goal of their project to be to answer the questions asked to them by newcomers. Specifically, this indicates that the Production activity that project members engage in will be on the Teahouse pages, answering questions from newcomers. From the perspective of the newcomer, however, this would equate to

either Social production or Social coordination, depending on the question. The choice of dependency in this case would be determined by the stated Sender and Recipient of any given action.

03\_0202 A: *And then in a very real way, because **I'm the guy that runs the bot that brings people in**, I have... I mean, some... my code is public, and the repo is listed from the, from the bot page.*

03\_0203 A: *But, but, it would be, it would be difficult if I were to just say "I'm not gonna do this anymore".*

03\_0204 A: *It would be... **it would take somebody else who's willing to put in a good deal of effort** to get those invites running again.*

03\_0205 A: *And without the invites, newcomer participation slows to a crawl.*

03\_0206 A: *Although there are plenty of other links to the Teahouse throughout Wikipedia.*

03\_0207 A: ***You really need to just spam people to get critical mass of new editors there.***

03\_0209 A: ***But, it's, but... no, like, no official hierarchy... and it's not, like, listed anywhere that I'm, you know... The host list is just dynamic, you know, based on level of participation.***

03\_0210 A: *My profile is there with everybody else's, and because I don't edit the page very much, my profile is seldom at the top.*

03\_0211 A: *So... it's really... it's pretty... there is a, there's a lot, there's **there's definitely an ingroup and an outgroup, but there's very few visible signals of what that ingroup and outgroup is, or, and very little defined hierarchy within that ingroup.***

03\_0212 Q: *Well, it sounds like the Teahouse, more than a lot of projects, has already enacted a bunch of, kind of technical solutions, socio-technical solutions to make sure that the project is lively, that it's active, and that people continue to go to it.*

03\_0213 Q: *I'm kind of curious, is there one part of that that seems like it may have not succeeded or not gone well? Or is there a direction that you wanna push this, like, what are the kind of the pluses and the minuses as you see them? And in kind of more of a.*

03\_0214 Q: *you know, "if anything was possible" sort of way?.*

...

03\_0219 A: ***It's hard to get new shit implemented on Wikipedia.***

## APPENDIX G: VIRTUAL TEAM EXPLORER – WIKIPROJECT SCENARIO SURVEY

### Introduction

The following survey intends to address the question, "How can we best facilitate tool-mediated coordination in distributed online teams?" To answer this question, we are going to present you with three separate scenarios. For each scenario, we will begin with a description of the context, the motivation behind the situation, and the steps that a Wikipedian takes to progress towards their stated goals. For each of these steps, you will be shown how that Wikipedian gathers information and acts on that information through the lens of a prototype tool, to aid in both the continued development and evaluation of that tool.

The survey should take about 20 minutes to complete. Once completed, answers will be stored on secured servers for further analysis by the research team. No data will ever be made publicly available in raw form, and your identity will never be identifiable connected with any data, even in aggregate. For completing this survey, you will be given a \$25 Amazon gift card.

If you have any questions about this survey, please contact Michael Gilbert at [mdg@uw.edu](mailto:mdg@uw.edu), or by phone at 206-354-3741. If you are ready to begin, please click Next below.

### Wikipedian #1 – Onboarding: Finding something to work on.

In this scenario, we are assuming the perspective of someone who is relatively new to Wikipedia - a new editor named Cameron.



*Source: Wikimedia Commons - Photographer.jpg*

Like many people, Cameron has known about Wikipedia for quite some time, but he only recently started looking into how to edit pages for a class project. By checking out some of the online documentation, he now feels like he knows the basics for how to edit articles, but he's still not sure about what he wants to edit.

In the process of going through some of the Getting Started documentation, Cameron comes across a link to a tool intended to help Wikipedians engage in collaborative group work - the Virtual Team Explorer, or the VTE. Curious,

he clicks on the link and follows the instructions to set the tool up for his account. Once he reloads the page, he sees a new link come up beside the search bar with the simple label: "VTE".



"Well that's pretty straightforward," he thinks. Clicking the VTE link, the application opens up as an overlay above the Wikipedia article he is currently reading.

### Scenario 1-1: Finding an active WikiProject

**Virtual Team Explorer** 👤 ⚙️ ✕

Enter a WikiProject to explore

Search for a WikiProject in the box above, or select from the list of most active WikiProjects below to continue.  
(Projects below represent the most active WikiProjects by edits to WikiProject pages and sub-pages within the last month)

Sort by edits
Sort by pages
Sort by ratio
Sort by WikiProject Edits

<p><b>WikiProject Good articles</b></p> <p>WikiProject Edits 645 Page Edits 74,121 Total Pages Edited 13,151 Edits per page 5.64</p>	<p><b>WikiProject Biography</b></p> <p>WikiProject Edits 465 Page Edits 48,919 Total Pages Edited 27,251 Edits per page 1.8</p>	<p><b>WikiProject Wikify</b></p> <p>WikiProject Edits 0 Page Edits 35,907 Total Pages Edited 10,190 Edits per page 3.52</p>
<p><b>WikiProject Featured articles</b></p> <p>WikiProject Edits 0 Page Edits 28,091 Total Pages Edited 3,731 Edits per page 7.53</p>	<p><b>WikiProject New Zealand</b></p> <p>WikiProject Edits 8 Page Edits 25,918 Total Pages Edited 24,417 Edits per page 1.06</p>	<p><b>WikiProject United States</b></p> <p>WikiProject Edits 349 Page Edits 17,758 Total Pages Edited 7,181 Edits per page 2.47</p>
<p><b>WikiProject Football</b></p> <p>WikiProject Edits 293 Page Edits 8,187 Total Pages Edited 5,671 Edits per page 1.44</p>	<p><b>WikiProject India</b></p> <p>WikiProject Edits 76 Page Edits 7,505 Total Pages Edited 3,616 Edits per page 2.08</p>	<p><b>WikiProject Wikipedia-Books</b></p> <p>WikiProject Edits 0 Page Edits 7,074 Total Pages Edited 2,555 Edits per page 2.77</p>
<p><b>WikiProject Australia</b></p> <p>WikiProject Edits 217 Page Edits 7,066 Total Pages Edited 4,177 Edits per page 1.69</p>	<p><b>WikiProject Politics</b></p> <p>WikiProject Edits 105 Page Edits 6,843 Total Pages Edited 2,086 Edits per page 3.28</p>	<p><b>WikiProject Insects</b></p> <p>WikiProject Edits 4 Page Edits 6,372 Total Pages Edited 5,427 Edits per page 1.17</p>
<p><b>WikiProject Former countries</b></p> <p>WikiProject Edits 9 Page Edits 5,514 Total Pages Edited 1,600 Edits per page 3.45</p>	<p><b>WikiProject Albums</b></p> <p>WikiProject Edits 120 Page Edits 5,472 Total Pages Edited 1,100 Edits per page 4.97</p>	<p><b>WikiProject Television</b></p> <p>WikiProject Edits 117 Page Edits 5,246 Total Pages Edited 1,100 Edits per page 4.50</p>

"Interesting," he thinks. Right off the bat Cameron can see the most active WikiProjects for the last month, by edits to the WikiProject pages themselves, edits to pages under the scope of those projects, the total number of pages under the scope of the project edited, and by the ratio of edits per page (within the scope of the project, the number of edits relative to the total number of articles edited). Given this ratio, it's fairly easy to identify projects that are seeing a lot of centralization in editing activity, or lots of edits to only a small number of pages.

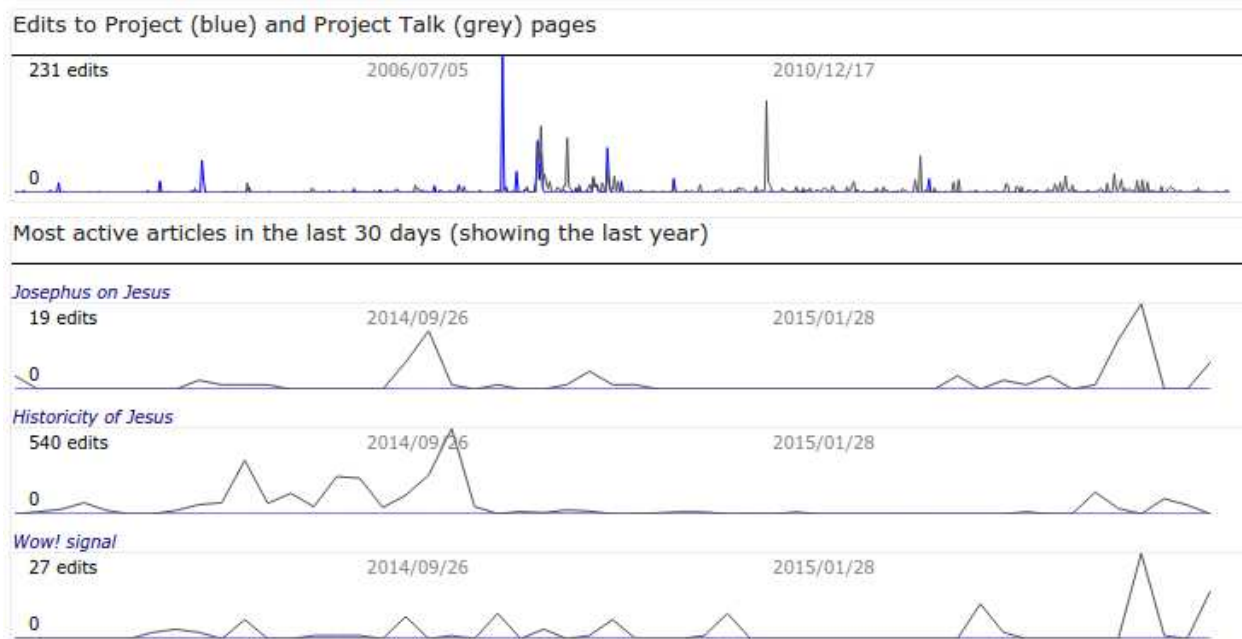
Cameron has always been interested in history, but the subject is so broad he's not sure where to start. To get an idea of his options, Cameron types "History" into the search box. The list of active projects automatically updates to those containing the phrase, "history." Because he wants to find an active project, or at least one where he's more likely to get a response if he posts on the Talk page, he sorts the results by WikiProject edits. This updated the top projects to those shown below:

<p><b>WikiProject Military history</b></p> <p>WikiProject Edits 227 Page Edits 302 Total Pages Edited 95 Edits per page 3.18</p>	<p><b>WikiProject History Merge</b></p> <p>WikiProject Edits 83 Page Edits 129 Total Pages Edited 25 Edits per page 5.16</p>	<p><b>WikiProject History</b></p> <p>WikiProject Edits 64 Page Edits 1,545 Total Pages Edited 488 Edits per page 3.17</p>
<p><b>WikiProject United States History</b></p> <p>WikiProject Edits 5 Page Edits 386 Total Pages Edited 130 Edits per page 2.97</p>	<p><b>WikiProject History of Science</b></p> <p>WikiProject Edits 2 Page Edits 799 Total Pages Edited 152 Edits per page 5.26</p>	<p><b>WikiProject Australian history</b></p> <p>WikiProject Edits 2 Page Edits 60 Total Pages Edited 14 Edits per page 4.29</p>

Looking through the updated list of projects, Cameron notices that active projects drop off pretty quickly. After the top 4 projects, sorted by edits to WikiProject pages and sub-pages, following WikiProjects all have two or fewer edits to project pages. "Well that's not what I want," he thinks.

## Scenario 1-2: Exploring an active WikiProject

Since WikiProject History was number three on the list, Cameron decides to select that project for a closer look. After clicking the project thumbnail, he's taken to a view of some of the high level project details, with the default view showing edits to the primary Project and Project talk page above a series of line charts showing a few of articles under the scope of the project which were most active over the last month -



Once again, he notices right away that, while it looks like activity on the main project page and talk page has slowed down, there's still a fair amount of activity on pages under the scope of the project. And, given the longevity of the project itself, it seems like it will probably still stick around for some time to come.

Interested, he looks at some of the most active articles under the scope of the project. Josephus on Jesus and Historicity of Jesus both seem promising but the third entry, Wow! signal, piqued his curiosity.

### Scenario 1-3: Finding an article to work on - Diving in

Cameron clicks on the link to take him to the article page, then takes a look at the Talk page to see what the recent editing activity was all about. Right at the bottom of the page, he finds the answer -

#### GA Review [\[edit\]](#)

*This review is transcluded from Talk:Wow! signal/GA1. The edit link for this section can be used to add comments to the review.*

**Reviewer:** [Jaguar](#) (talk · contribs) 18:24, 27 May 2015 (UTC)

I should get this done by tomorrow **JAGUAR** 18:24, 27 May 2015 (UTC)

#### Initial comments [\[edit\]](#)

- "Amazed at how closely the signal" - informal, try "impressed"?
- I would recommend expanding the lead slightly (a line or two maybe) in order to summarise, and also splitting in the lead into two paragraphs would be great
- "when the 37-year old Ehman spotted a surprising vertical column" - is it necessary to include his age in this sentence? Seems like it was extracted from somewhere
- "The circled alphanumeric code **6EQUJ5**" - per **WP:BOLD** it isn't required to embolden anything other than the lead title. This would be better in quotes
- The first half of the Time variation section is unreferenced
- I would recommend merging some of the smaller paragraphs in the Searches for recurrence of the signal section to improve prose flow
- The See also section is a bit long, I think maybe cutting down a couple of links would suffice (there is a guideline on too many external links somewhere, but I can't remember it!)


#### References [\[edit\]](#)

- [Ref 12](#) is dead

#### GA toolbox

- [Peer review](#)
- [Dup detector](#)
- [Disambig links](#)
- [External links](#)

The article was being reviewed! "Well I can handle some of these," he thinks. "Here are some clearly stated needs that I can just tackle, one by one!" Cameron, now armed with both the knowledge to find active WikiProjects as well as some tricks to identify the most active articles under the scope of that project, starts fixing up the page given the review completed by the user, Jaguar. Once he's all done, he posts an update on the talk page asking if everything looked good. A bit later he got a response -

Thanks for addressing them! Looks good now  **JAGUAR** 18:24, 28 May 2015 (UTC)

Success! "Now," Cameron thinks, "That wasn't too tough. I wonder what it takes to review an article..."

#### Questions:

- 1) Given the scenario above, how useful do you think a tool like this would be for newcomers to quickly find active WikiProjects (with 1 being least useful, 4 being most useful)?

- 2) How likely would you be to use a tool like the VTE, described above (with 1 being least likely, 4 being most likely)?
- 3) The scenario above involves a relatively new editor to Wikipedia, searching for an active group of Wikipedians to more effectively identify both clear and manageable tasks that he could complete, as well as a larger group of editors that he could more actively conspire with. What kinds of additions or modifications would you make to a tool such as the VTE that might make achieving these goals easier for Cameron? What kinds of changes or modifications might make achieving these goals easier for yourself?

## **Wikipedian #2 - Task coordinator: Organizing the commons to help the team get stuff done**

In this scenario, we are assuming the perspective of a seasoned WikiProject member seeking to better organize the work that needs to get done in a project, the work that's already been completed, and the editors that are engaging in that work. Meet Charlie - hobbyist scientist, avid cook, and experienced Wikipedian.



*Source: Wikimedia Commons - Woman-typing-on-laptop.jpg*

Charlie has been a long-time Wikipedian but recently decided to join a relatively new project, WikiProject Women writers, that she felt had an important place in the community but may need a boost to make sure members were able to more easily identify and complete work that might benefit the project. To make things simpler, Charlie wanted to focus on making a few different processes easier for project members: categorizing articles, reaching out to new members, and tracking the work that project members were completing more effectively.

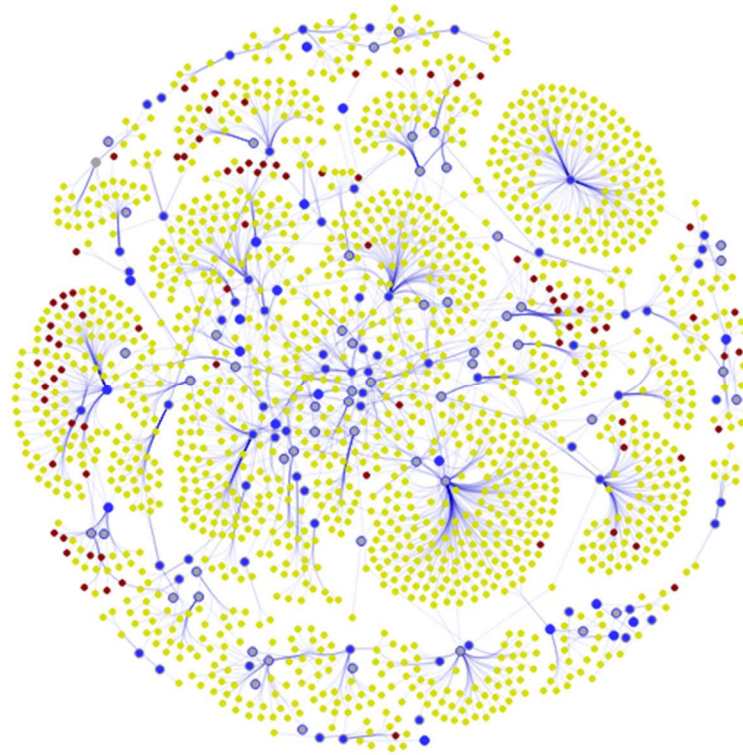
When Charlie first found the VTE, she noted that the documentation stated that the tool was intended to help WikiProjects structure and standardize some of the more common tasks that project members endeavor to complete. "Well that sounds right up my alley," she thought, and installed the User Script by adding a single line to a special Wikipedia page. As soon as that was done, she started digging into what this little tool could do.

### **Scenario 2-1: Categorizing articles**

As an experienced Wikipedian, Charlie knows that there are a number of different ways that articles can be categorized. Frequently, this involves project members identifying articles that may fit in the category through exploration or experience, and then manually adding a link to the category on the article page. While Charlie knows there are some tools out there which can make this task easier - HotCat, for instance, allows editors to add, modify,

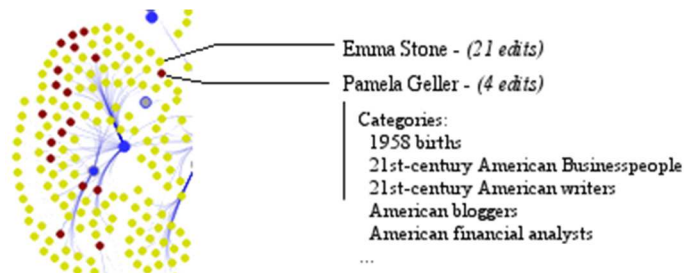
or remove categories in just a click - the task of first identifying those potential articles can sometimes be problematic.

Using the VTE, Charlie is able to visualize a network graph originating from every Wikipedian who recently interacted on the WikiProject page, talk page, or any sub-pages for the project -



In this graph, Charlie can see project members (blue nodes), other Wikipedians who are not project members (grey nodes), along with the articles that they have each edited recently - with red nodes indicating an article inside the scope of the WikiProject (i.e., they are already categorized under the WikiProject) and yellow nodes indicating an article currently outside the scope of the WikiProject.

Charlie immediately notices that there are many articles being edited by WikiProject members that aren't currently categorized as under the scope of the project. With a quick glance, she notices one user near the left of the graph who appears to have been fairly actively editing many articles, both in- and out-side the project's scope -

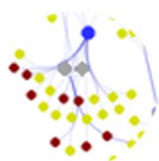


By mousing over the yellow nodes (the articles currently not under the scope of the WikiProject), Charlie can quickly see a few details on the article, including the article title, how actively it is being modified (indicated by both the number of edits and the number of unique editors to that article and its talk page), as well as any projects that article is currently claimed by. In just a few clicks, Charlie can then choose to visit the article to determine whether it should be categorized as under the scope of WikiProject Women writers or she can simply add the category link from right within the VTE.

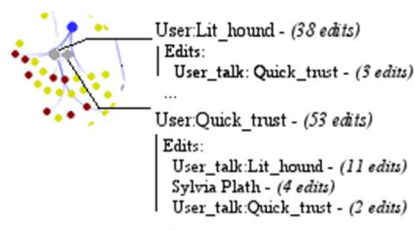
Ultimately, starting from any Wikipedian who has recently interacted on a WikiProject Women writers related page, Charlie is able to get a high level view of what those Wikipedians are working on and how it might impact the project itself.

## Scenario 2-2: Reaching out to potential WikiProject recruits

Another common task that project members go through is making sure there is a critical mass of active editors to ensure the project doesn't fizzle up or become forgotten - while it's possible to have a very successful project with only half a dozen members, that's usually only the case when the majority of those members are active. To make sure that WikiProject Women writers has the critical mass it needs to succeed, Charlie also wants to send out some invitations to the more active non-project members to see if they want to join. To do this, she once again turns to the network visualization above. Immediately she sees a few users that she might want to look into, but she focuses on two in the bottom-left of the graph -



"That's curious," she thinks to herself. "It looks like there are two non-project members working together with a project member, on a mix of articles both in- and out-side the scope of the project."



By focusing on that part of the graph, Charlie is able to quickly focus on the types of work these three editors were working on, including the amount of activity they are dedicating to those articles, and the amount of communication that's occurring on the talk pages for each article.

"It looks like these two have been working together for some time," Charlie notices. "I think I'll send them a quick invitation to the project and see if they might want to work with the larger group. With another click, Charlie brings up a menu for each user that includes the option to send the user an invitation to the project. Clicking that option,

Charlie then customizes the default invitation text to be a bit more familiar and hits the Send button which, all from within the VTE, constructs a new talk page section and posts it to each users' talk page.

"Well, that should do it!" she exclaims. "I'd bet I'll be hearing from them soon."

*Questions:*

- 4) Given the scenario above, how useful do you think a tool like this would be for Wikipedians to manage common WikiProject-related tasks (with 1 being least useful, 4 being most useful)?
- 5) How likely would you be to use a tool like the VTE, described above (with 1 being the least likely, 4 being the most likely)?
- 6) The scenarios described above involve a fairly experienced editor utilizing a new tool to both structure WikiProject-related tasks as well as to make completing some of those tasks easier. What kinds of additions or modifications to this tool might you suggest that might make Charlie's goals of organizing WikiProject Women writers easier? What kinds of changes or modifications might make achieving these goals easier for yourself?

### **Scenario 3-1: Creating tasks for group members, and tracking how they're progressing**

Finally, Charlie wants to check on the status of some of the outstanding task requests that the project has on its Tasks page. For many projects, this would be a fairly simple list of tasks, potentially linking to individual articles that need to be assessed, updated, copyedited, or otherwise fixed up. But Charlie knows the VTE is capable of creating and maintaining these types of lists, in a way that seamlessly allows both human editors (i.e., the group members) as well as the VTE User Script to update and interact with the same content.

What's special about this, though, is that the specific way this is implemented allows the VTE to include any data it needs within those tasks. This "extra" data can be easily ignored when it is displayed within Wikipedia if it is overwhelming, as it is a Module invocation which decides what gets displayed and what doesn't, but it remains accessible to the VTE (as well as any other tools, of course), that choose to interact with it. What this means, Charlie soon found out, is that any list of tasks or requests (or, really, any other list at all) can have virtually any type of data associated with it.

In the case of the task list that Charlie was working with, this included the task requestor, the creation date, a title, a more detailed description, the estimated time it would take to complete the task, as well as any arbitrary number of subtasks that may be associated with or required to complete the parent task. In-wiki, this would look like -

Title	Description	Priority
Style check	Could someone make sure that <a href="#">Web design</a> conforms to desired style?	4
Request for mentor	I'd like someone to let me know what types of articles may be best for a newbie to work on	5
Vandal watch for <a href="#">Mobile interaction</a>	This article has been vandalized a few times in the last week, need help monitoring	0

Where in the VTE, this task list may look like -

+ Add task		View: All ▼	Sort: Created ▼
Apr 7, 2015	4	Style check	4 comments
Apr 4, 2015	5	Request for mentor	0 comments
Apr 1, 2015	0	Vandal watch for <a href="#">Mobile interaction</a>	0 comments

In this latter view, any of the above information can be included in the default columns depending on the information Charlie is curious about. Tasks that are "overdue" can be colorized, and conversations about a task (taken from a Talk page section created when the task is first created) can be easily identified and browsed through. Viewing the task details by clicking on the task allows Charlie to update any information for each task she might want - including adding new subtasks, updating the time to completion, or adding any relevant notes or questions that may be pertinent to completing the request, all in one place.

To finish up her editing session, Charlie looks for any new tasks created since the last time she logged in and, finding only one old task that was recently marked as closed by the VTE due to not being automatically triggered (the third vandal watch task above), decides it's time to wrap up for the night.

### Questions:

- 7) Given the above scenario, how useful do you think a tool like this would be to track and control common tasks in Wikipedia (with 1 being the least useful, 4 being the most)?
- 8) In closing, each of the scenarios described above are intended to illustrate how a unique tool can integrate various channels of information to facilitate how WikiProject members can organize and complete their work. Extending from this goal, do you have any comments or criticisms regarding the general goal, or ideas that you feel may be more effective in supporting WikiProject members?

## Concluding confirmation

Thank you for completing the survey! Your participation greatly helps as we work towards identifying ways to best support coordination and participation within WikiProjects. I will issue you the gift card as soon as responses are received.

Lastly, if you are curious about the tools described in this survey, the current prototype version of the VTE and its documentation are available at [https://en.wikipedia.org/wiki/User:Md\\_gilbert/vte](https://en.wikipedia.org/wiki/User:Md_gilbert/vte). We'll be continuing development on this tool taking your feedback into consideration and hope to be able to release a beta version in the coming months.

Once again, thank you, and please feel free to contact me if you have any questions.

-Michael Gilbert, [mdg@uw.edu](mailto:mdg@uw.edu)

## APPENDIX H: IDENTIFYING AUTOMATED TOOL AND BOT EDITS

As referenced in Part II, Study 3, “Non-bot Edits Identified Through Standard Strings in Revision Comments,” tool-mediated edits were initially identified utilizing a list of comment strings for common tools originally published on the Wikipedia Toolserver. However, as the Toolserver was decommissioned in June, 2014 (along with the page that hosted these comments), I am including the revision comments as an Appendix to this dissertation to ensure the reproducibility of the analyses described above.

```
UPDATE rev_table SET tool = 'rollback' WHERE rev_comment RLIKE "(Reverted ([0-9]+
)?edits by \\[[Special:Contributions/[^\\|]+\\][^]]+\\) \\[[User
talk:[^\\|]+\\|talk\\]\\]) to last version by .+>";
UPDATE rev_table SET tool = 'undo' WHERE rev_comment LIKE "Undid revision%";
UPDATE rev_table SET tool = 'huggle' WHERE rev_comment LIKE '%WP:HG%' OR rev_comment
LIKE '%WP:HUGGLE%';
UPDATE rev_table SET tool = 'huggle' WHERE rev_comment RLIKE "(Message re\\.
\\[[^]]+\\)\\)|(Level [0-9]+ warning re\\. \\[[^]]+\\)";
UPDATE rev_table SET tool = 'twinkle' WHERE rev_comment LIKE '%WP:TW%';
UPDATE rev_table SET tool = 'friendly' WHERE rev_comment LIKE '%WP:FRIENDLY%' OR
rev_comment LIKE '%WP:Friendly%';
UPDATE rev_table SET tool = 'vandalproof' WHERE rev_comment LIKE '%WP:VPRF%' OR
rev_comment LIKE '%WP:VandalProof%' OR rev_comment LIKE
'%VandalProof|VandalProof%' OR rev_comment LIKE '%WP:VP2%' OR rev_comment LIKE
'%WP:VandalProof%';
UPDATE rev_table SET tool = 'stiki' WHERE rev_comment LIKE '%|STiki]]%';
UPDATE rev_table SET tool = 'npwatcher' WHERE rev_comment LIKE '%|NPWatcher%';
UPDATE rev_table SET tool = 'vandalsniper' WHERE rev_comment LIKE '%|VandalSniper%';
UPDATE rev_table SET tool = 'wikimonitor' WHERE rev_comment LIKE '%m:WikiMonitor%';
UPDATE rev_table SET tool = 'mwt' WHERE rev_comment LIKE '%MWT|MWT]]%';
UPDATE rev_table SET tool = 'awb' WHERE rev_comment LIKE '%AWB|AWB]]%' OR rev_comment
LIKE '%AutoWikiBrowser%';
UPDATE rev_table SET tool = 'cluebot' WHERE LOWER(CAST(rev_user_text AS CHAR)) LOWER
LIKE 'cluebot%';
UPDATE rev_table SET tool = 'antivandalbot' WHERE rev_user_text = 'AntiVandalBot';
UPDATE rev_table SET tool = 'orphanbot' WHERE rev_user_text = 'OrphanBot';
UPDATE rev_table SET tool = 'pseudobot' WHERE rev_user_text = 'PseudoBot';
UPDATE rev_table SET tool = 'voabot' WHERE rev_user_text LIKE 'VoABot%';
UPDATE rev_table SET tool = 'martinbot' WHERE rev_user_text = 'MartinBot';
UPDATE rev_table SET tool = 'stbot' WHERE rev_user_text LIKE 'STBot%';
UPDATE rev_table SET tool = 'sqlbot' WHERE rev_user_text = 'SQLbot';
UPDATE rev_table SET tool = 'tawkerbot' WHERE rev_user_text LIKE 'Tawkerbot%';
UPDATE rev_table SET tool = 'sinebot' WHERE rev_user_text = 'SineBot';
UPDATE rev_table SET tool = 'csdwarnbot' WHERE rev_user_text = 'CSDWarnBot';
UPDATE rev_table SET tool = 'antispanbot' WHERE rev_user_text = 'AntiSpamBot';
```

## **APPENDIX I: TOWARDS A CLASS-BASED MODEL OF INFORMATION ORGANIZATION IN WIKIPEDIA: DATA STRUCTURES UTILIZING SEMI-STRUCTURED TEXT**

As complexity increases in commons-based peer production communities, the means of organizing and facilitating collective action must also mature to ensure the ongoing health and active maintenance of those communities (Marwell & Oliver, 1993). This study examines the types of structured data that exist in Wikipedia, introduces an argument for an extension to the types of structured and semi-structured data within Wikipedia supported by that descriptive analysis; and presents an implementation of that extension that supports instantiations of semi-structured content that facilitate both human and tool-mediated interactions with Wikipedia data. This extension offers a novel means of structuring data to support the ongoing health and maintenance of online communities like the community of editors that maintain and develop Wikipedia.

### **Situating Proposed Structures among Current Practice**

Because Wikipedia has grown for more than a decade, supporting the goal of representing the “sum of all human knowledge” requires an increasingly complex ecosystem of tools and processes to ensure that community members can effectively create, update, and maintain quality content across the myriad categories of knowledge now encompassed in the encyclopedia. These tools range from the simple to the complex, from built-in templates and modules that form part of the core of the MediaWiki platform, to sophisticated agents such as ClueBot NG, an autonomous bot user that utilizes advanced machine learning to identify and automatically revert vandalism in the encyclopedia. And while the user experience for most visitors to the encyclopedia is limited to a rather seamless display of encyclopedic content, that content and the community responsible for its creation rely on many tools, a web of functionality that fundamentally facilitates and alters both how users experience Wikipedia and how editors contribute to it. For example, one of the most frequently used tools, the Yesno template, is responsible for normalizing an input to be a yes/no output, and has been utilized over 6.7 million times<sup>70</sup>. ClueBot NG, a vandalism-fighting bot, made over 600,000 edits to the English language Wikipedia in 2013 alone. Further, of the ten most active accounts on Wikipedia during 2013, eight were bot accounts and the remaining two were primarily tool-mediated edits – namely, edits made with the ProveIt<sup>71</sup> and HotCat<sup>72</sup> tools, neither of which are part of the MediaWiki platform, but which facilitate and simplify interactions with the content stored within it.

Clearly, interactions within Wikipedia are increasingly affected by the tools available to the community, both those that are enabled by default such as the templates and modules so frequently used and the many other tools that exist outside the platform to aid in editing efforts. While such tools enable many varied interactions with the Wikipedia content and community, significant opportunities to improve information organization and presentation within the encyclopedia are numerous. For example, the system would benefit from an improved, robust means of organizing

---

<sup>70</sup> Wikimedia Tool Labs, *supra* note 9.

<sup>71</sup> [https://en.wikipedia.org/wiki/User:ProveIt\\_GT](https://en.wikipedia.org/wiki/User:ProveIt_GT), allowing editors to more easily “find, edit, add, and cite references in Wikipedia articles.”

<sup>72</sup> <https://en.wikipedia.org/wiki/Wikipedia:HotCat>, allowing editors to easily add or remove categories from individual or groups of articles with a single click.

information. Such an enhancement could provide a powerful and extensible way to extend the efforts of individual editors. This study presents a principled approach to the design of such a tool, using an analysis of existing tools and solving for a problem that has not yet been addressed by a system tool.

Our approach introduces a tool capable of storing and presenting tabular lists of arbitrary data in a way that is resilient to both human and non-human edits in the encyclopedia. This approach allows for storing information about an arbitrary object (the data) as well as for defining relevant behavior for the validation and presentation of that object (the code). By adopting an object-oriented language to define and present this work, our tool-based approach is distinctly different from existing mechanisms in Wikipedia.

The goals of this study are three-fold. First, we conduct a quantitative analysis of existing uses of templates and modules in Wikipedia, concluding with a brief descriptive analysis of those most frequently in use. Second, we discuss the benefits and shortcomings of current approaches, highlighting a direction to move forward with more resilient strategies. And, finally, we introduce a new method of data storage that allows for a robust and stable means of defining object information in Wikipedia. Our tool, employing this method, will allow editors to gain more control over tabular data and its presentation within the system. Our approach emphasizes the need for the data to be fault tolerant, to avoid breaks by human edits, as well as to be flexible, facilitating multiple methods of presentation between multiple destinations. Our new data structure thus utilizes existing potential in the MediaWiki platform to allow more nuanced control of information storage and presentation within Wikipedia.

## **Related Work: Prior Explorations of Tool-Mediated Coordination and Information Organization**

To guide our work, we sought to first understand the potential behind structured and semi-structured data in Wikipedia. Additionally, we considered the use of automated and semi-automated tools to mediate the ongoing health and operation of Wikipedia. As we discovered, the intersection between these two classes of related work indicates the need for a novel approach to supporting human efforts and tool-mediated interactions.

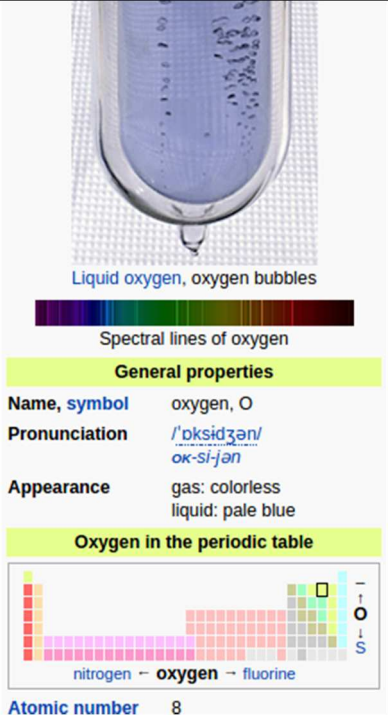
In several studies, researchers have considered different means of both structuring semantic data in Wikipedia as well as means of mining that data for external use. Bizer et al. (2009) explore how the DBpedia project extracts information for infoboxes in Wikipedia, archives them locally, and makes these data pairs accessible to third party applications via a public API on the World Wide Web. These infoboxes represent a common means of data re-use in Wikipedia. Using these infoboxes, editors have a simple and straightforward means of defining any number of attribute-value pairs and the presentation of that information can be delegated to a specialized Template. See Figure 21 below for an example of both the declaration and instantiation of one such infobox for the chemical element Oxygen. Beyond the simple instantiation, presentation, and extraction of attribute-value pairs present in these infobox templates (e.g., Wu & Weld, 2010), another project (Lange, Böhm, & Naumann, 2010) explores means of automatically populating empty attribute pairs in Wikipedia articles, building on prior work by Wu & Weld (2008).

Related to these examinations of semi-structured data as information object is prior research on mediating tools in Wikipedia. This “bespoke code” that supports the on-going operation of the encyclopedia makes up an estimated six million lines of code, an order of magnitude greater than the about 600,000 lines of code that represent the core

MediaWiki platform (Geiger, 2014). These tools are crucial for the ongoing health of the encyclopedia, providing editors with the means to patrol and revert vandalism, monitor edits, and increase the visibility of information intended to aid editors in their ongoing efforts, mediating the work that occurs between the “social” and the “technical” (Ribes et al., 2013).

Our contribution presents a variation of bespoke code that allows editors to interact with the system in ways that have thus far not been possible. Beyond that, by creating a novel means of structuring tabular data in Wikipedia we provide a new avenue for research on semantic data extraction and re-use to progress (such as Bizer et al., 2009; Wu & Weld, 2010; Wu & Weld, 2008) providing a vast amount of contextual data that can be mined and interpreted, with the potential to inform the design and implementation of new tools in the future.

In the next section we lay out a more detailed description of the templates and modules used within Wikipedia, including their construction, prevalence and utility, as well as potential shortcomings with regard to managing more complex data types requiring greater flexibility in both presentation and behavior. Following, we introduce a novel approach to utilizing these existing affordances to resolve these shortcomings and to better support complex cooperative work.

<pre> {{infobox element  name=oxygen  number=8  symbol=O  pronounce={{IPAc-en ˈ ɒ k s i dʒ ə n}}  pronounce 2={{respell OK si-jən}}&lt;!-- no need for US vs. UK. They're the same thing.--&gt;  left=[[nitrogen]]  right=[[fluorine]]  above=  below=[[sulfur S]]  series=diatomic nonmetal  group=16  period=2  block=p  series color=  appearance=gas: colorless&lt;br&gt;liquid: pale blue  image name=Liquid Oxygen.gif  image alt=A glass bottle half-filled with a bluish bubbling liquid  image size=170  image name comment= [[Liquid oxygen]], oxygen bubbles  image name 2=Oxygen spectre.jpg ... </pre>	 <p>The image on the right is a composite of several elements related to oxygen. At the top is a photograph of a glass bottle containing a pale blue liquid with small bubbles rising from it, labeled "Liquid oxygen, oxygen bubbles". Below this is a spectral line chart for oxygen, showing a series of colored lines (violet, blue, green, yellow, red) labeled "Spectral lines of oxygen". Underneath the chart is a table titled "General properties" with the following entries: Name, symbol: oxygen, O; Pronunciation: /'ɒksɪdʒən/ (with a link to the pronunciation) and ɒk-si-jən; Appearance: gas: colorless, liquid: pale blue. Below the table is a periodic table titled "Oxygen in the periodic table" where oxygen is highlighted in a yellow box. The periodic table shows elements from hydrogen to fluorine, with labels "nitrogen", "oxygen", and "fluorine" at the bottom. At the very bottom of the composite image, it says "Atomic number 8".</p>
---	--

**Figure 21.** An example of the instantiation and presentation of the infobox template for the element Oxygen. The cell on the left contains the text used to define the template attributes, the image on the right is part of the resulting presentation.

## The Current Landscape

### Templates and Modules – Introduction

Many tools built directly into the MediaWiki platform simplify the process of creating, formatting, and maintaining different types of information. As introduced above, primary among these are templates and modules. The idea behind these tools is simple: once created, both templates and modules allow editors to define simple key value pairs, and once instantiated the template or module will handle the presentation of that data. As shown in Figure 21 above, these template instantiations can be arbitrarily complex, allowing the template designer to craft complicated visualizations from the information passed in by the template caller (separating the instance of the template, or its instantiation, from the class, its definition), although more frequently templates are used for “boilerplate messages, standard warnings or notices, infoboxes, navigational boxes, and similar purposes.”<sup>73</sup> An instance of a template can be created on any Wikipedia page by simply including the template name in braces. For example, a simple template transclusion may be to include “{{lambda}}” on a page, which would result in the lambda symbol being included in the resulting text, “λ.”

Modules are similar to templates in both their instantiation and capacity to simplify the structure and presentation of data – many modules have actually been created to replace existing templates. However, while templates define their presentation through standard Wiki markup, modules allow users to define custom Lua scripts that are run when the module is transcluded on the target page. This allows users a more powerful means to “analyze data, calculate expressions, and format results using functions or object-oriented programming.”<sup>74</sup> The ability to run these Lua scripts on the MediaWiki platform was added through the Scribunto interface, originally installed on the English language Wikipedia in February, 2013.

### Templates and Modules in Use

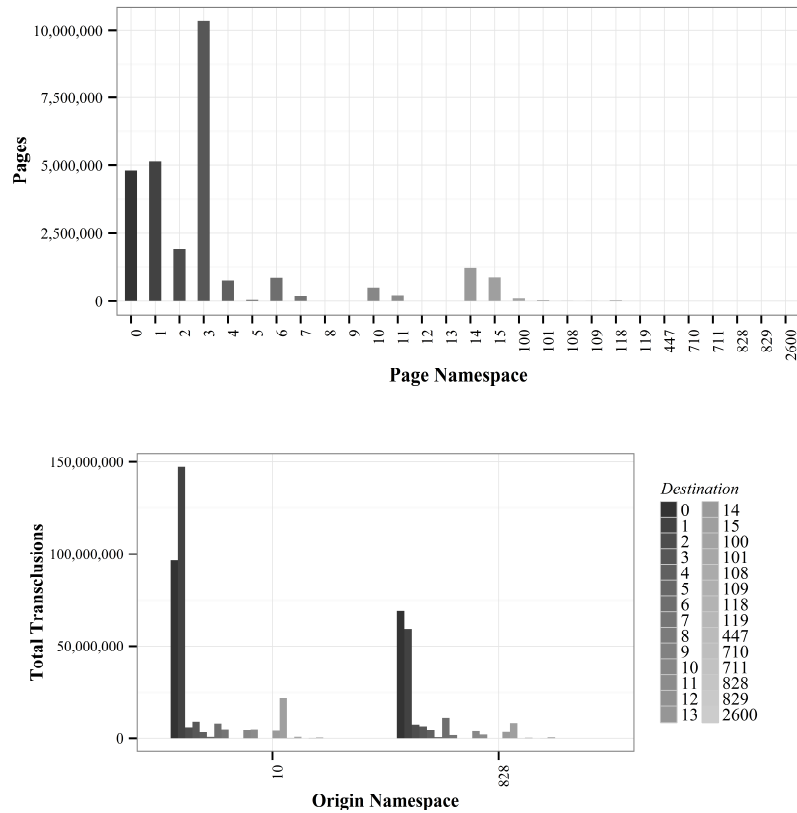
The first template was checked into the English language Wikipedia on September 30, 2002. Since that time, templates have become one of the most ubiquitous and frequently used tools available to Wikipedians, with over 605,000 templates created, transcluded over 312 million times on pages within the English Wikipedia. Modules, with their more recent release date, currently only number just over 2,100. However, despite their more modest number, modules have currently been transcluded over 179 million times on pages. As the total number of pages across all namespaces<sup>75</sup> (including the template and module namespaces) is just over 35 million, it is evident that many of these transclusions are occurring multiple times per page, highlighting a pattern of usage verifying the central role that these tools play in the presentation of standard information within Wikipedia. For a more detailed view of the distribution of pages over namespaces, as well as the distribution of template and module transclusions over namespaces, see Figure 22.

---

<sup>73</sup> <https://en.wikipedia.org/wiki/Help:Template>

<sup>74</sup> <https://en.wikipedia.org/wiki/Wikipedia:Lua>

<sup>75</sup> Namespaces are how the MediaWiki software categorizes pages of distinct type, for instance allowing the clear separation between Article pages and User pages. See <https://en.wikipedia.org/wiki/Wikipedia:Namespace> for more detail.



**Figure 22. (a) Distribution of pages across namespaces, (b) Template and Module transclusions on all pages, split by destination namespace.**

The ubiquity of these tools is clear, but what are they most often being used for? The single most used Template is the Yesno Template referred to above, which is currently transcluded over 6.7 million times. Following that is the Transclude Template, with over 6.1 million transclusions, which handles resolving page names to clarify Wikipedia links. Top among Modules is Arguments, with over 15 million transclusions, a meta-module which enables other Modules to process arguments passed in when they are invoked. Following that is the Module version of the Yesno Template, with over 11 million transclusions. The top templates and modules that are currently in use in the English Wikipedia, along with the number of transclusions for each, is included in Table 13 below.

**Table 13. The most frequently transcluded Templates and Modules in the English Language Wikipedia.**

Template	Transclusions	Module	Transclusions
Yesno	6,763,997	Arguments	15,526,763
Transclude	6,197,441	Yesno	11,481,409
Navbar	6,182,642	No_globals	7,192,709
Pagetype	6,107,802	Namespace_detect	6,230,069
WPBannerMeta	6,049,255	Pagetype	6,107,727
Class_mask	6,003,990	Category_handler	5,759,067

## Existing Shortcomings

While these templates and modules exist for a wide array of purposes, addressing both simple and complex needs, they are not without their shortcomings. First, our primary motivator to embark upon this analysis was driven by a need to have a means of modifying semi-structured text by both human and non-human agents to more effectively support tool-mediated modifications of tabular data within Wikipedia. In this context, semi-structured text must be machine parsable in a predictable format, allow for any amount of attribute-value pairs within that format, and finally, support values that can be an open text string of any length. This flexibility was required to ensure that a single class of module could be instantiated with multiple data sets, facilitating a broad set of data manipulation and presentation possibilities. Second, our use case required granular control of error handling, so that if human agents entered malformed or unexpected data in the instantiation of the module that it could attempt to correct or properly display the remaining data. And third, we desired a solution that had the capacity to normalize input data, as well as providing a mechanism for Wikipedians to have a broad range of control over the style of presentation of that data.

Finally, while solutions like infobox templates work adequately for defining more complex information about individual items, such as the Oxygen molecule in Figure 21 above, we required a solution that would provide a means of defining lists of arbitrary semi-structured data. While it is trivial to instantiate a table using Wiki markup, if that table does not contain predictable data encoded in a predictable fashion, it is a more difficult task to utilize a tool-mediated approach to interact with that data. Consider for instance Wikipedia's "List of astronauts by name,"<sup>76</sup> presented as a simple bullet-point list of astronauts, sorted alphabetically, including various additional information about each one such as former names, birth and death dates, and the shuttles each astronaut went up on. Creating a tool that could read this list, parse it, and update it would be fairly simple. However, attempting to utilize that same tool to update the many other lists on the English Wikipedia<sup>77</sup>, as well as the many other task lists, member lists, and list of lists present on WikiProject pages, would likely fail. Without a structured way of storing and presenting this vast assortment of semi-structured text, the number of tools required to support tool-mediated edits to that text could conceivably approach the total number of lists in Wikipedia. Accordingly, a more flexible approach is needed to ensure that this data can be accurately parsed by both human and non-human agents, to more effectively support both the individuals who create and consume that content, as well as the researchers who could far more easily mine the semantic data present in such lists if it were accessible in a more viable format.

No such structure currently exists. In the sections that follow, we lay out the strategies we adopted to create this data structure, and introduce our initial implementation of that approach.

## Improving Information Design

In defining our implementation of the data structure defined above we had two primary goals. First, we wanted to address the shortcomings identified to ensure that our approach would be resilient to potentially incomplete or malformed data, enabling both tool-mediated and manual modifications. And second, we wanted our solution to

---

<sup>76</sup> [https://en.wikipedia.org/wiki/List\\_of\\_astronauts\\_by\\_name](https://en.wikipedia.org/wiki/List_of_astronauts_by_name)

<sup>77</sup> According to the Wikimedia Tool Labs database, there are 187,959 pages in the Article namespace in the English Wikipedia with a title starting with "List of," indicating that this could be an approach to parsing semantic data in semi-structured text with wide-reaching implications.

utilize an approach that is already highly accepted and widely adopted by the Wikipedia community; namely, we wanted to adopt the module as the primary means of storing and presenting data. This implementation allowed us to increase the usability of our approach by reducing the complexity of adoption, as well as ensured that we could take advantage of the expanded functionality offered by the Lua scripting language compared to existing templates.

## Templates and Modules Expanded

For our implementation, we created two separate modules. The first is a parent module that manages parsing the arguments list, determining the style of the resulting presentation, and handling errors or malformed data. The primary argument to this parent module is a series of calls to a child module, which are treated as the data items that should be presented as the output of module instantiation. Specifically, when instantiating our implementation of a list module, four steps are taken. First, the parent module determines the style of presentation. This could take the form of a standard table, a page section, or any repeatable form of presentation that MediaWiki is capable of displaying. Second, the parent module determines the elements that should be displayed in Wiki markup, required to ensure that parameters passed as individual data items are correctly normalized and accounted for before calling the child module. This includes ensuring that the data items that are defined within the parent module invocation have the required values to be displayed properly defined in their sub-module invocation or that a suitable default exists, as well as ensuring that malformed data passed to the parent module is ignored so that the remaining data can be properly parsed and displayed. The actual data in the child module inclusions does not matter – the parent module can determine the greatest common set of identical attributes and present the data as requested. And fourth, the child module invocation strings are created and processed, returning Wiki markup for the desired format for each data item presented to the parent module. An initial implementation of these modules is available from Wikipedia under `Module:ListMaster` and `Module:ListItem`.

To take a more specific example, recall the list “List of astronauts by name” article mentioned above, which included alphabetized bullet points for each astronaut, potentially but not always including additional information such as alternate names, birth and death dates, and the name of the shuttles the astronauts were on. Utilizing the approach described above, this list could be implemented similar to the following (this example was broken across multiple lines to increase readability, although this is not required):

```
{ {#invoke:ListMaster|printTable|style=bullet|
  display=name,lifespan,shuttle#|
  { {#name=[[Joseph M. Acaba]]|
    shuttle0=[[STS-119]]|shuttle1=[[Soyuz TMA-04M]]
  }}
  { {#name=[[Loren Acton]]|
    shuttle0=[[STS-51-F]]
  }}
  { {#name=[[Mike Adams]]|
    lifespan=(1930-1967)|shuttle0=[[X-15]]
  }}
  { {#name=[[James Adamson]]|
    shuttle0=[[STS-28]]|shuttle1=[[STS-43]]
  }}
}}
```

By providing a simple, fault tolerant way to structure arbitrary semi-structured data, editors have the capacity to flexibly control the style and structure of their data. For example, by modifying the invocation above from `style=bullet` to `style=table`, each astronaut would be represented as a row on a table, sortable by any column, with the built-in ability to provide reasonable defaults for missing values and to skip or attempt to correct malformed data items, here defined as the sub-module invocations inside the `#invoke` statement. Further, due to the structured nature of the semi-structured data, it becomes a trivial matter to parse attribute value pairs from any arbitrary list defined in Wikipedia, providing a unique means to collate and analyze data as well as to create novel tool-mediated interactions that have the capacity to further shape and improve the way that we interact within that online community.

## **CONCLUSION**

The approach presented in this study is not meant to be a monolithic solution to replace module usage, but as a means to improve editors' abilities to manage their own content in a manner that is both human editable and machine parsable, extending the research listed above by allowing simple constructs to inform the display of arbitrary objects within the community. By defining an extensible solution for storing, validating, and displaying data within Wikipedia, we hope to facilitate a more flexible means of information presentation than is now possible, as well as providing new means for researchers to parse and utilize an identifiable instantiation of the wide array of semi-structured text present within Wikipedia.