

**Collective Organization for Public Good Provision:
The Case of Wikipedia**

by

Luis Fernando Heredia

A Thesis Submitted in Partial Fulfillment of the Requirements for the Honours Degree
of Bachelor of Art in the Department of Economics

Supervised by:

Dr. Pascal Courty

We accepted this extended essay as conforming to the required standard

Dr. Daniel Rondeau, Honours Co-Advisor (Department of Economics)

Dr. Herbert Schuetze, Honours Co-Advisor (Department of Economics)

Luis Fernando Heredia

4/25/2012

University of Victoria

Collective Organization for Public Good Provision:

The Case of Wikipedia¹

Wikipedia is an online-encyclopedia, based on an openly editable model and written entirely by volunteers without pay. Since its creation in 2001 Wikipedia has grown enormously, receiving over 7 billion page views in January 2012. Having amassed over 3 million content pages in the English Wikipedia alone, it is today the largest repository of human knowledge (Wikipedia, 2012). Its unprecedented success and the manner in which this success has been achieved pose interesting challenges to our traditional understanding of the closely related theories of public goods and collective-action.

These theories predict a socially suboptimal outcome resulting from the inherent non-rivalrous and non-excludable nature of public goods. Inefficiency due to non-rivalry results from individuals' failure to take into account the marginal cost imposed on society by their actions, and is usually equated with a negative externality caused by the over-use or production of a given resource or commodity (Layard, 1978). Upon closer scrutiny, this externality arises when the quantity of individuals consuming a particular good grows so as to increase the degree of rivalry such as congestion on roads or individuals' over-use and eventual degradation of a natural resource.

Inefficiency due to non-excludability arises from the inability to extract all rents resulting from a given action (Sandmo, 1987; Layard, 1978). In this case, individuals will have an incentive to hide their true preferences and contribute little to production activity in the hope that others may gain sufficiently from the provision of such a good so as to engage in its production. In either case, inefficiencies may be resolved through the creation of institutions that adequately align private and public costs and benefits; institutions that will not come about voluntarily as

¹ Special thanks are provided to Drs. Courty, Rondeau, and Schuetze for their guidance and patience in the development of this thesis. All mistakes are my own.

individuals will free-ride by disguising their preferences so as to enjoy the benefits brought about by others actions (Sandmo, 1987). Furthermore, assuming an institution were to be created, there is no guarantee that individuals will impose sanctions on others when they are incurring a personal cost while benefits are widely distributed (Ostrom, 1990). In other words, there is a lack of incentives that ensure credible commitment to the design principles agreed upon by individuals. These sources of inefficiency have been found to be especially pronounced in cases of open access or communal ownership (Alchian, 1987). However, evidence suggests that individuals often overcome the social dilemmas of collective-action (Ostrom, 1990). Case studies in countries such as Spain, Italy, Japan, and the Netherlands have found that individuals create institutions for sustainably managing communally owned natural resources (Ostrom, 1990; Ostrom, 2003).

Wikipedia is challenging in that it has embraced the very characteristics public good and collective-action theories have determined to be causes of failure. Why would thousands of individuals freely contribute to the provision of a public good? Given individuals ability to enjoy access to its content, there is no a priori reason to believe individuals would contribute to Wikipedia if they believed others would. Individuals may enjoy the benefits of other content providers while not incurring the costs of engaging in content production. Furthermore, allowing anyone to edit its pages exposes Wikipedia to over-use in that the quality of its content may be sub-optimal. Assuming individuals were to contribute, it is likely that a larger quantity of content providers will lower the quality of Wikipedia's content due to higher coordination costs and a decrease in the likelihood of reaching consensus.

Despite these potential shortcomings, Wikipedia has flourished. With anonymous users and 80,000 active contributors, making over 11 million edits a month to 19 million pages in over 270 languages, Wikipedia can today be looked at as a massive experiment in collective action (stats.wikimedia.org; Viegas, Wattenberg, Kriss & van Ham, 2007). Furthermore, studies have found its content to be stylistically similar to traditional printed sources, in terms of formality

and language standardization while the accuracy of its content has been found to be the same as that of Britannica articles (Emigh & Herring as cited in Viegas et al, 2007).

Open-source collaboration has also experienced growing success (von Hippel & Von Krogh, 2002). Similar to Wikipedia, open-source collaboration relies on individual user innovations that are then provided as public goods (Benkler, 2002). Through loosely defined property rights, individuals are able to modify a program to suit their particular needs, resulting in a larger variety of options for others to draw from. These forms of collaboration, also defined as commons-based peer production or institutions for collective-action, are increasingly becoming a significant economic and social phenomenon (Benkler, 2002; von Hippel & Von Krogh, 2002; Lerner & Tirole, 2002).

The increasing application of these kinds of institutions to the production and provision of various kinds of public goods provides a strong motivation for their further investigation. Institutions such as MoveOn.org, the Khan Academy, and Wikipedia facilitate political agency and knowledge creation by granting individuals the ability to voice their concerns and share knowledge. NASA Clickworkers relies on public volunteers to do routine science analysis that would normally be done by a scientist or graduate student (Benkler, 2002). Prosper, an on-line peer-to-peer bank, allows individuals to personally choose how much and to who they wish to lend. This has led scholars to argue that open-source collaboration may possess certain advantages to traditional property- and contract-based forms of production (Benkler, 2002; von Hippel & von Krogh, 2002). A greater understanding of the social benefits derived from collective-action institutions, their various applications, and the characteristics that contribute to their success would increase the tools and courses of action policy makers and individuals may have at their disposal when deciding between various kinds of socio-institutional arrangements.

Overview

This research contributes to a burgeoning literature that attempts to identify the sources of Wikipedia's strengths. I provide a nuanced description of the institutional similarities between open-source collaboration and successful common-pool resource institutions enshrined in Wikipedia's institutional functioning and operations. I provide an important contribution to the literature by attempting to statistically determine the extent to which WikiProjects have contributed to Wikipedia articles quality by decreasing Wikipedians monitoring and coordination costs. I employ data from various Wikipedia pages' revision histories and their respective WikiProjects quality assessments to conduct a cross-sectional regression analysis. Its main finding is that WikiProjects play an important role in facilitating coordination activity that impedes reductions in the quality of its top tier articles while their impacts on lower quality articles remain uncertain.

Section 1 begins by drawing on the insights provided by the open-source economics literature to describe the motivations of individual contributions and the advantages conferred to Wikipedia by granting all users the right to modify its content. Sections 2 & 3 describe Wikipedias' institutional structure and its similarities with the design principles of successful common-pool resource (CPR) institutions. Section 4 attempts to determine the benefits conferred to Wikipedia by the application of CPR design principles to the process of content creation. Section 5 describes the data and methodology used to discern the extent to which WikiProjects have contributed to higher quality content on Wikipedia. Section 6 presents and discusses results and potential limitations. To conclude I briefly expand on the results of this investigation and the problems of coordination, rather than underproduction, Wikipedia is faced with.

1) Some Simple Economics of Wikipedia

Research regarding the micro-motivations underlying open-source collaboration have pointed to the signaling incentives of career concerns, institutional identification, and ego gratification or peer-recognition as the main drivers of free software development (Lerner and

Tirole, 2002; Lerner, Pathak and Tirole 2006; Anthony et al 2009; Forte et al, 2009). These signaling incentives may take the form of potential career opportunities or enhanced reputation resulting from high-quality contributions to a project or page, inherent value derived from designing or modifying a program to better suit a developers specific needs, or the furthering of practical skills that come about from doing so (Lerner & Tirole, 2002; Anthony et al, 2007). These benefits counterbalance the opportunity costs of time incurred by collaborators that may manifest themselves as forgone earnings or decreased performance in other areas (Lerner et al, 2006). Developing a proper understanding of the economic value that results from socially productive behavior has led scholars in this area to argue that open-source collaboration presents advantages over other forms of organization in reducing the degree of information loss in organizing economic activity and the transactions costs of cooperation (Benkler, 2002).

In order for individuals in collaborative production systems to be productive, institutions must design ways of providing individuals with signals that are used to choose between various courses of action given their talents and preferences (Benkler, 2002). To perform these functions, markets and firms rely on prices and institutional hierarchies to determine the most efficient use and combination of agents, efforts, and resources. The process of specification results in a degree of information loss due to the inability to properly define and match human and material resource characteristics (Benkler, 2002). Open-source collaboration more adequately identifies various kinds of human creativity, and allocates it to its most efficient use by placing the decision of what project or page to contribute to with the individual. Doing so provides a framework for individuals who possess the best knowledge about their fit for various tasks to choose what tasks to perform (Benkler, 2002). Open-source collaboration also substantially reduces transaction costs by removing property and contract as organizing principles which limit individuals' ability to scour larger groups of materials and collaborations (Benkler, 2002).

There are substantial increasing returns to granting large clusters of individuals' access to large repositories of information in search of new projects. When individuals are allowed to contribute as much or as little as they please, the incentives required to motivate individual

contributions are reduced (Benkler, 2002). This facilitates the pooling of people's various levels of effort and capabilities and ensures that private benefits are roughly proportional to the costs of contributing. The costs of monitoring and correcting malicious or unwanted edits in Wikipedia, coupled with the benefits of granting free-access, suggests that benefits from anonymous contributions will be inversely proportional to the quantity of existing information on a given content page.

The size of Wikipedia articles is limited by various technical issues and the quantity of information required to adequately cover a given topic. Momentarily setting aside heterogeneity in the quality of individuals contributions and assuming that quality is solely determined by a topics coverage, an anonymous contribution to a stub (an article deemed too short to provide encyclopedic content) poses a much smaller threat to an article's quality than it would to a featured article, Wikipedia's top article-quality rating. The probability of an anonymous contribution positively contributing to a page's quality is inversely related to the size of a particular page. This follows from the fact that the "opportunity cost of quality" increases the larger the quantity of information. As quality rises, the marginal benefit from allowing everyone to contribute will increasingly depend on Wikipedian's ability to accurately differentiate between positive and negative contributions. Their ability to do so will in turn depend on providing an institutional context that reduces the costs of monitoring, exclusion, reaching consensus, and correcting unwanted edits.

The size of Wikipedias' content pages exhibit characteristics similar to those of club goods which are defined as being non-rivalrous but excludable. In these settings increasing the size of a group may confer additional resources that could be used to provide a benefit that will be enjoyed by all such as greater content creation (Ostrom, 2003). On the other hand more individuals will also lower the marginal cost of providing a good to all beneficiaries (Ostrom, 2003). The benefits accrued to Wikipedia by anonymous user contributions will be a function of the costs of monitoring, exclusion, reaching consensus, and correcting unwanted edits. These

costs will ultimately depend on the community generated social norms that are articulated in Wikipedia's artifacts of governance.

2) Wikipedia as Collective Organization for Public Good Provision

In its beginnings Wikipedia policy-making was informal and done on a trial and error basis (Forte et al, 2009). As the project grew and coordination costs increased, it became harder to reach consensus. This led to the pseudo formalization of the policy-making process, the rise of a complex institutional hierarchy, and the development of tools to facilitate the control of the content creation process (Forte et al, 2009).

a) Policies, Rules, and Norms:

Wikipedia has developed policies and guidelines that specify acceptable conduct and editing activity. Examples include its emphasis on the importance of a neutral point of view (NPOV), which provides individuals with strong motivations to avoid bias and conflict. This policy clarifies the kind of content that may be presented on a page by requiring its verifiability through the use of cited sources, although this does not guarantee inclusion. The NPOV policy has been found to be frequently referred to by Wikipedians when resolving disputes regarding agreement on the inclusion of diverse subject matter (Butler, Joyce & Pike, 2008; Viegas, Wattenberg & Dave, 2004). Wiki policy tends to reflect community practices by being normative rather than positive; a principle epitomized in its "Wikipedia does not have firm rules" policy (Wikipedia: Five pillars; Butler et al, 2008). Finally, Wikipedia stresses respectful and civil interaction between editors through policies such as "assume good faith" by emphasizing the importance of reaching consensus, and following dispute resolution procedures (Wikipedia: Five Pillars; Butler et al, 2008).

b) Institutional Hierarchies:

Individuals will often choose to become involved in specialized tasks, such as reviewing

articles at others' request, watching current edits for vandalism, or watching newly created articles for quality control purposes. Wikipedians may build a reputation as competent editors and take on various roles subject to peer approval. Editors who believe they can serve the community better by taking on additional administrative responsibility may ask their peers for agreement to undertake such responsibilities. This structure enforces meritocracy, communal standards of editorship and conduct, and tends to ensure a high level of experience, trust, and familiarity across a broad range of aspects within Wikipedia.

Previous analyses of the various social roles that make up the Wikipedia community have identified three user states that fundamentally affect the kind of social authority individuals may exert. Unregistered or anonymous users have little influence in shaping policies or norms, although they are an important part of the institution's identity by holding the ability to freely edit the encyclopedia (Forte, Larco & Bruckman, 2009). Registered users, also known as Wikipedians, can play a wide variety of roles, hold nuanced forms of authority, and be granted certain privileges. Like Wikipedia, CPR institutions, or institutions for collective-action, rely on a broader conception of property rights than the one traditionally employed in much of the economics literature (Ostrom, 2003).

c) Various forms of Property-regimes:

Scholars studying the characteristics of successful CPR institutions have identified five types of property rights that are usually relied upon for the sustainable management of natural resources. These are defined as follows:

Access: the right to enter a defined physical area and enjoy non-subtractive benefits (e.g. hike, canoe)

Withdrawal: the right to obtain resource units or products of a resource system (e.g. catch fish, divert water).

Management: the right to regulate internal use patterns and transform the resource by making improvements.

Exclusion: the right to determine who will have an access right, and how that right may be transferred.

Alienation: the right to sell or lease exclusion, management, or withdrawal rights.

While property-rights systems that do not include the right to alienation are traditionally considered ill defined, research has shown that CPR institutions often use different property-rights systems along side each other (Ostrom, 2003). Individuals may then take advantage of economies of scale in cases where increasing the size of a group may confer additional resources that could be used to provide a benefit that will be enjoyed by all, while also lowering the private marginal cost of providing a good to all beneficiaries (Ostrom, 2003). This implies a convex production function defined by positive interdependence that has been shown to be a characteristic of club goods (Ostrom, 1997). Consistent findings across many studies have found that individuals develop communal property regimes to take advantage of this fact (Ostrom, 2003). For example, long-enduring communally owned irrigation systems frequently allocate water and maintenance costs using the amount of land owned by a farmer (Ostrom, 2003). Striking similarities between CPR institutions and Wikipedias' reliance on various forms of property rights are summarized in the following two tables:

Table 1: Access Levels in English Wikipedia

Administrator	Protect/unprotect pages; Delete/undelete pages; Block/unblock users; Special revert tools.
Bureaucrat	Make administrators; Rename users; Make other bureaucrats.
Steward	Change all user access levels on all wikimedia projects.
Oversight	Hide page revisions from all other user types.
Checkuser	View user IP addresses.
Developer	Access to MediaWiki software and Foundation servers (various sublevels).

Source: Forte, Larco, & Bruckman, 2009

Table 2: Bundles of Rights Associated with Positions

	Full Owner	Proprietor	Authorized Claimant	Authorized Entrant	Authorized Entrant
Access	X	X	X	X	X
Withdrawal	X	X	X	X	
Management	X	X	X		
Exclusion	X	X			
Alienation	X				

Source: Ostrom, 2003

Wikipedia has granted anonymous users the rights to access and withdrawal while registered users may possess differing degrees of rights to management, exclusion, and alienation. The various tasks undertaken by Wikipedians and the rights assigned to those tasks will be strongly associated with the incentives they most respond to. Reliance on various forms of property rights takes advantage of the economies of scale that result from allowing individuals to determine the quantity of effort they wish to put towards the provision of a public good. In addition to the institutional hierarchies that have arisen to facilitate the creation of content, Wikipedians have developed various tools to reduce the costs of monitoring, enforcement, exclusion, and coordination.

3) Decentralizing Bureaucracy:

Wikis allow for the archiving of all previous edits and make it simple to revert to an earlier version (Viegas et al, 2004). The archived versions of a page, or a page's history, allow users to view the differences between saved versions, the date and time when a change occurred, who made the change, either a user name or IP address in the case of an anonymous user, and any comments made by the contributor (Viegas et al, 2004). Wikipedia has also developed ways for users to discuss potential changes to articles content, monitor the changes that are made to specific pages, and restrict contributions from anonymous individuals.

The majority of pages on Wikipedia have a corresponding talk page. Although talk pages are intended for discussing improvements to a particular article, Viegas et al (2007) found that Wikipedians frequently rely on talk pages to engage in the strategic planning of editing activity, reach consensus, and facilitate the maintenance of articles. Talk pages have been shown to play a crucial role in granting users the ability to articulate what they deem to be the main issues to resolve for the improvement of an article's quality (Viegas et al, 2007). Wikipedians also rely on Watchlists to notify them of any changes that are made to a page. This allows committed contributors to closely scrutinize the edits of suspicious anonymous users. Though not often relied on, various levels of edit and move protection may be implemented to impede certain users

from contributing to a page. Just as policies and guidelines evolved during Wikipedia's early growth to specify modes of conduct and content creation, Wikiprojects have arisen to reduce higher coordination costs resulting from further growth (Forte et al, 2009).

Wikiprojects are decentralized social structures composed of a coalition of editors with common goals or expertise (Forte et al, 2009). They are commonly organized around subject matter within which local leadership, norms, and loosely defined stylistic guidelines are established (Forte et al, 2009). For example, the United States Wikiproject is composed of over 1,200 active members comprising 66 other Wikiprojects. Wikiprojects may themselves be composed of other Wikiprojects, departments, or task forces used to coordinate activity for specific areas of need (Forte et al, 2009). This allows Wikipedians to focus on articles in particular areas and develop policies specific to an articles needs.

Wikiproject's jurisdiction may overlap with others in cases where a page's content is related. The Cold War page for example, falls under the purview of thirteen Wikiprojects. No governance mechanisms currently exist in situations such as these or where local policy must be enforced. The standard process of discussion and consensus building is usually relied upon to resolve local disputes frequently referencing Wikipedia's broader guidelines, only some of which were mentioned above (Viegas et al, 2003; Forte et al, 2009). Wikiprojects policies are therefore nested within the organization. In cases where consensus is not reached, the parties involved may resort to administrators. If this were to fail, individuals may resort to an arbitration committee that serves as Wikipedia's Supreme Court. The arbitration committee is charged with the interpretation of policy and the enforcement of sanctions in cases where previous attempts at dispute resolution by administrators or users themselves have failed (Forte et al, 2009). Scholars investigating the causes of Wikipedia's success have pointed to similarities between its institutional organization and the factors shown to result in successful long-enduring CPR institutions (Viegas et al, 2007; Forte et al, 2009).

The following table has been adapted from Ostrom (1990) to reflect Forte et al's (2009) discussion regarding the application of the design principles illustrated by successful CPR institutions to Wikipedias' case and is meant to provide a brief depiction of the similarities between these kinds of institutions.

Table 3: Similarities between design principles illustrated by long-enduring CPR institutions and Wikipedias' institutional functioning:

CPR Institutions	Wikipedia
<p>1. <i>Clearly defined boundaries</i> Individuals/households who have rights to withdraw resource units must be clearly defined, as must the boundaries of the CPR.</p>	<p>1. Wikiprojects and registered and anonymous users.</p>
<p>2. <i>Congruence between appropriation and provision rules and local conditions</i> Appropriation rules restricting time, place, technology and/or quantity of resource units are related to local conditions and to provision rules requiring labour, material, and/or money.</p>	<p>2. Wikiproject members' ability to devise stylistic guidelines specifically attuned to an article's particular needs.</p>
<p>3. <i>Collective choice arrangements</i> Those affected by operational rules can participate in modifying them.</p>	<p>3. Wikiprojects, Talk Pages and the assigning of privileges based on merit and experience.</p>
<p>4. <i>Monitoring</i> Monitors are accountable to the appropriators or are the appropriators.</p>	<p>4. Administrators</p>
<p>5. <i>Graduated sanctions</i> Appropriators who violate operational rules are likely to be assessed graduated sanctions by other appropriators, by officials accountable to these appropriators, or by both.</p>	<p>5. Warnings and Bans</p>
<p>6. <i>Conflict-resolution mechanisms</i> Appropriators and their officials have rapid access to low-cost local arenas to resolve conflicts among appropriators or between appropriators and officials.</p>	<p>6. Talk Pages, administrators, and access to the Arbitration Committee.</p>
<p>7. <i>Minimal recognition of rights to organize</i> The rights of appropriators to devise their own institutions are not challenged by external governmental authorities.</p>	<p>7. Wikiproject members' ability to devise stylistic guidelines specifically attuned to an article's particular needs.</p>
<p>8. <i>Nested enterprises</i> Appropriation, provision, monitoring, enforcement, conflict resolution, and governance activities are organized in multiple layers of nested enterprises.</p>	<p>8. Wikiprojects, their respective departments/task forces and their role within the wider Wikipedia community.</p>

Research conducted by Viegas et al (2007) regarding the importance of talk pages to coordinate activity and reach consensus suggests that the quantity of registered contributions to a talk page may have a positive impact on an articles quality. This may be due to efficiency gains in the allocation of effort that result from individual cooperation more accurately identifying areas in need of improvement. This could also be due to a decrease in the amount of conflict resulting from providing Wikipedians with a forum to resolve their differences. However, it is not clear whether anonymous contributions to a talk page will have a positive or negative impact.

Previous research by Anthony et al (2009) has shown that the proportion of a contributions retained characters tends to be higher for registered users. However, they also found that anonymous users with low levels of participation often provide contributions with a character retention rate that is higher than that of registered users while controlling for article and contribution size. These findings coupled with the discussion of benefits derived from open-source collaboration in section 2 suggest that anonymous retention rates will be inversely related to a pages quality and size. As mentioned in Section 3 the use of Watchlists may also contribute to higher quality content by decreasing monitoring costs. The benefits from monitoring activity will be especially important in the case of high-quality articles where the probability of an anonymous user making a positive contribution becomes increasingly low. Finally, Wikiprojects may contribute to an article's quality by further reducing the costs of monitoring, enforcement, and cooperation.

4) Fishing for Institutional Causality:

The application of these design principles alone is not enough to guarantee the success of collective-action. Examples of successful and unsuccessful attempts by governments, cooperatives, voluntary associations, and private individuals to govern and manage open-access resources abound (Ostrom, 2003). While a variety of research regarding Wikipedias' institutional development and functioning as well as Wikipedians changing roles has been undertaken,

attempts to determine the impacts these have had on the provision of quality content have been limited despite this being Wikipedia's most important goal. Kittur and Krout (2008) and Anthony et al (2009) seem to be the only ones who have attempted to determine the causes of high quality pages. However, neither of them focuses on the extent to which Wikipedia's institutional design affects the quality of its content. Despite the benefits derived from the application of these design principles to the management of CPRs, the impacts they have had on Wikipedia remain unclear. Keeping constant the quantity of anonymous and registered users there is no reason to believe that Wikipedia's institutional characteristics would result in higher quality content due to greater levels of effort units or their more efficient allocation.

Assuming an arbitrary average quality for anonymous contributors that is below that of registered users, a page's quality will depend on the quantity of effort units allocated to monitoring and exclusion or production activity. Holding registered users constant, an increase (decrease) in the quantity of anonymous contributions will decrease (increase) quality while the opposite will result from an increase in the quantity of registered users if we assume a fixed quantity of anonymous contributions. Furthermore, when keeping both of these constant as well as production activity, a decrease (increase) in the cost of monitoring and exclusion costs will increase (decrease) the amount of effort units that may then be assigned to the production of content. However it is not clear that this would be the case. In the absence of strong incentives individuals could choose to allocate newly liberated effort units to leisure activity. The same could be said to occur in the case of an increase in the marginal impact of a productive unit of effort on quality. This could result from a Wikiproject's creation and development of stylistic guidelines. Again however, the absence of strong incentives could lead individuals experiencing marginal productivity gains to decrease the effort put towards productive activity resulting in an equal or lower quantity of content production. It is to answering these puzzles that we now turn.

5) Data & Methods:

The limited amount of research regarding the impacts of Wikipedia's institutional characteristics on the quality of its content allows for the investigation of a range of potentially significant variables.

a) Data:

To determine the impacts of Wikiprojects on article quality I use cross-sectional data from various Wikipedia pages revision histories ranging from January 1st, 2009 - April 5th, 2012. The sample consists of twenty pages from each quality class resulting in a total sample size of 240 pages comprised by 120 articles and their corresponding talk pages. The fact that the sample is random within but not across quality categories should not be a cause for concern as the purpose of this analysis is not to arrive at inferences regarding the composition of a population's specific characteristics but to determine the impacts of an observable characteristic that is prevalent throughout the population. The gathering of data required the creation of a program that would query Wikipedia for the revision histories of the specified articles and their respective talk pages. A database was then created to store this data from which it was later extracted and formatted through the use of SQL queries.

The data includes variables such as the total amount of edits and reverts, broken down into those made by registered and anonymous users and the quantity of Wikiprojects involved in an article. Further data regarding the amount of individuals "watching" a page, and the average number of views received by an article was gathered manually from the historical statistics provided by Wikipedia. The average number of views was computed using the quantity of views received by each page during the last month of each of the four periods for which data was collected. In following Kittur and Krout (2008) I take advantage of articles' respective Wikiproject quality assessments by assigning them a value from 1-6, with higher numbers being associated with higher quality classes. Although Wikiprojects may create their own metrics for the determination of an article's quality, the general criteria provided by Wikipedia to determine an article's quality ratings is summarized in Table 4 below.

Quality Index	Table 4: Quality Assessment Criteria	
6	Featured Article (FA)	Professional, outstanding, and thorough; a definitive source of encyclopedic information.
5	A	Very useful to readers. A fairly complete treatment of the subject. A non-expert in the subject matter would typically find nothing wanting.
4	Good Article (GA)	Useful to nearly all readers, with no obvious problems; approaching (although not equalling) the quality of a professional encyclopedia.
3	B	Readers are not left wanting, although content may not be complete enough to satisfy a serious student or researcher.
2	C	Useful to a casual reader, but would not provide a complete picture for even a moderately detailed study.
1	Start	Provides some meaningful content, but the majority of readers will need more.

There may be heterogeneity of quality ratings among Wikiprojects pertaining to a particular page. However, the variance between average quality and the highest quality assigned to a page by a given Wikiproject was minimal in this particular sample. Thus, the maximum quality assigned to a page was used to determine where a page might lie within the above distribution.

b) Hypotheses:

Previous research by Anthony et al (2009) has shown that the proportion of retained characters tends to be higher for registered users suggesting that anonymous contributions will on average have a negative impact on an article's quality. Research conducted by Viegas et al (2007) regarding the importance of talk pages to coordinate activity and reach consensus suggests that the quantity of contributions to a talk page may have a positive impact on an articles quality. This may be due to efficiency gains in the allocation of effort that result from individual cooperation more accurately identifying areas in need of improvement. This could also be due to a decrease in the amount of conflict resulting from providing Wikipedians with a forum to resolve their differences. However, it is not clear whether anonymous contributions to a talk page will have a positive or negative impact.

As mentioned in Section 3 the use of Watchlists could contribute to higher quality content by decreasing monitoring costs. The emphasis on the design principles of CPR institutions increasing the likelihood of their success suggests that Wikiprojects may play an important role in Wikipedias ability to increase the quality of its content. Wikiprojects could contribute to an article's quality by further reducing the costs of monitoring and reaching consensus which may result from the creation of page and topic specific stylistic guidelines. These hypotheses are supported by the descriptive statistics of these institutional variables presented in Tables 5 & 6 which splits the sample into two groups with good quality pages having a quality rating greater than or equal to four and low quality pages a rating lower than or equal to three.

Tables 5 & 6: High and Low Quality pages institutional characteristics.

High Quality	<i>Content Pages</i>					<i>Talk Pages</i>				
	<i>Mean</i>	<i>Median</i>	<i>Std Dev</i>	<i>Skewness</i>	<i>#Obs</i>	<i>Mean</i>	<i>Median</i>	<i>Std Dev</i>	<i>Skewness</i>	<i># Obs</i>
<i>Total Edits</i>	818.81	515.55	228.25	87.75	6183	139.216	42	207.15	2.52	8353
<i>Registered</i>	590.56	366.5	673.44	2.304	35434	121.01	39	179.92	2.528	7261
<i>Anonymous</i>	228.25	366.5	412.03	17.5	13695	18.18	3.5	32.08	2.33	1091
<i>Reverts</i>	87.75	673.447	152.83	2.62	756	4.95	0	11.80	3.89	297
<i>Wikiprojects</i>	3.47	3	2	.67	132	-	-	-	-	-
<i>Avg Views</i>	51552.38	10590.75	84682.94	2.24	3093142.5	-	-	-	-	-
<i>Watchers</i>	148.72	48.5	240.74	2	8923	-	-	-	-	-

Low Quality	<i>Content Pages</i>					<i>Talk Pages</i>				
	<i>Mean</i>	<i>Median</i>	<i>Std Dev</i>	<i>Skewness</i>	<i># Obs</i>	<i>Mean</i>	<i>Median</i>	<i>Std Dev</i>	<i>Skewness</i>	<i># Obs</i>
<i>Total Edits</i>	353.254	168	79	7.94	2706	32.66	12	2	8.87	271
<i>Registered</i>	215.20	118	14	2.15	1057	25.64	11	1	6.51	181
<i>Anonymous</i>	145.44	48.5	12	16.99	1649	5.91	1	0	21.71	90
<i>Reverts</i>	42.18	10	0	8.33	351	1.77	0	0	21.81	21.81
<i>Wikiprojects</i>	2.51	2	1.72	3.24	206	-	-	-	-	-
<i>Avg Views</i>	18729.49	4098.13	31761.22	2.60	1123769	-	-	-	-	-
<i>Watchers</i>	57.08	15	109.33	3.80	3425	-	-	-	-	-

As shown, high quality pages exhibit a larger proportion of registered to total edits. There is also a large disparity between the quantities of contributions made to high and low quality articles' talk pages. It is worth noting that the average quantity of reverts in both high and low quality pages is lower than the average quantity of anonymous contributions suggesting that anonymous contributions can increase and articles quality. This would be expected to depend on registered users ability and willingness to engage in monitoring and exclusion activity. High quality pages also tend to have a larger quantity of Watchers and Wikiprojects overseeing their editing process. Finally, Figure 1 presented below suggests that Wikiprojects may play a significant role in either increasing a pages quality or impeding its decrease.

Figure 1:



However, a simple comparison of means does little to confirm the previously mentioned hypotheses as these outcomes may be the result a variety of factors that are uncorrelated with the benefits resulting from the provision of efficient institutions. Among these we may include an articles age, exogenous real world events that may lead to a page receiving greater levels of attention, and issues of reverse causation that may result from the selection bias of high quality articles attracting editors rather than editing activity improving article quality (Kittur & Krout,

2008). This warrants the further investigation of the interplay between Wikipedians and the institutional context within which they operate.

b) Model & Limitations:

The correlational nature of cross-sectional data and a small sample size limit the conclusions that may be drawn from this investigation. Cross-sectional analyses do not adequately address issues of reverse causation or control for time invariant unobserved characteristics like an articles topic, importance, or the availability of secondary sources (Kittur & Krout, 2008). The use of ordinary least squares and the ordinal nature of the quality ratings I wish to investigate means that reported coefficients cannot be interpreted as the marginal impact resulting from a unit increase in an explanatory variable. They are thus left out of the results presented below. However, the small sample size allows for the inclusion of data regarding institutional variables that may only be gathered manually. Furthermore, given the lack of research surrounding the impacts of Wikipedia's institutional characteristics and the ambiguity of the effects these may have on the quality of its content, results may provide important areas for further research.

A linear probability model is used as a first attempt to determine the impacts of institutional characteristics on Wikipedias' ability to provide quality content. I rely on the use of a dummy variable that takes the values of 0 and 1 as my dependent variable. The dummy is initially assigned to pages within the top 2 quality classes, as it is assumed that institutional characteristics will have played an important role in their development. This assumption is then relaxed by adding GA quality pages to my dependent variable. The quantity of contributions made by registered and anonymous users, Wikiprojects involved in a page, and the quantity of individuals who have a given article on their Watchlist are used as explanatory variables. Given the quantity of edits by anonymous and registered users all have highly skewed distributions I follow Kittur & Krout (2008) by taking the natural log of these variables in order to partially normalize their distributions. As controls I rely on the average quantity of views received by a given page and the length of content and talk pages as proxies for an article's age and popularity.

6) Analysis & Results:

The following analysis proceeds by investigating the impacts of individual and institutional characteristics first in content and then in talk pages. In both cases, registered and anonymous users are treated separately.

a) Content Pages:

Results from the regression that uses articles with a quality rating above five as a dependent variable are consistent with previous findings. Registered user contributions are positively correlated to high quality articles and statistically significant at the 1 percent level. Wikiprojects are also statistically significant at the 1 percent level and positively correlated to quality, suggesting that institutional characteristics may play an important role in Wikipedia's ability to provide quality content. However, the significance of both registered user contributions and Wikiprojects disappear when controlling for a pages size and its average number of views. Anonymous contributions are not statistically significant in any of the specifications when not combined with registered users. When contributions by both kinds users are included in the regression anonymous contributions are negative and statistically significant at the 1 percent level and remain significant even after controlling for number of views and a pages length. Surprisingly the quantity of individuals monitoring a page is not statistically significant. Upon further investigation it was found that a significant positive correlation between the quantity of users "watching" a page disappears when either registered contributions or the quantity of Wikiprojects are included in the regression.

The more interesting case arises when the dependent variable is expanded to include articles with a GA quality rating, the results of which are presented in Tables 7 & 8 below. In this case the negative relationship between registered users reverts becomes highly statistically

significant even after controlling for views and size. One possibility is that too much monitoring activity and not enough coordination at lower quality levels is decreasing the gains that could result from retaining a greater quantity of anonymous contributions. The negative relationship between registered reverts and lower quality articles may be what precludes anonymous contributions from becoming statistically significant. Although registered user contributions remain positive and highly significant, Wikiprojects no longer have an impact on a page's quality. However, Wikiprojects remain highly significant and positive when included in the regression that investigates the impacts of anonymous users on content pages. This lends support to the proposition that monitoring and exclusion activity may be a disservice to Wikipedia when it comes to lower quality articles.

Table 7:
Impacts of Registered Users on Content Pages

Quality Dummy ≥ 4	1	2	3	4	5
<i>Log of Registered Edits</i>	(+) ^{***} (.0233)	(+) ^{***} (.0260)	(+) ^{***} (.0377)	(+) ^{***} (.0364)	(+) (.04358)
<i># of Wikiprojects</i>	-	(+) (.0217)	(+) (.0210)	(+) (.0232)	(+) (.0233)
<i>Log of Reg Deletions</i>	-	-	(-) ^{***} (.0299)	(-) ^{***} (.0319)	(-) ^{***} (.0295)
<i># of Watchers</i>	-	-	-	(+) (.0002)	(-) (.0005)
<i>Average Views</i>	-	-	-	-	(+) (.0000)
<i>Content Length</i>	-	-	-	-	(+) ^{***} (.0000)
R-Squared	0.165	0.190	0.240	0.241	0.338
# Obs	120	120	120	120	120

***p <.01, **p <.05, *p <.1

Table 8:
Impacts of Anonymous Users on Content Pages

Quality Dummy ≥ 4	1	2	3	4	5
<i>Log of Anonymous Edits</i>	(+) (.0228)	(+) (.0220)	(+) (.0208)	(-) (.0204)	(-) (.0235)
<i># of Wikiprojects</i>	-	(+) ^{***} (.0206)	(+) ^{***} (.0199)	(+) ^{***} (.0246)	(+) (.0249)
<i>Log of Anon Deletions</i>	-	-	(+) ^{***} (.215)	(-) (.1992)	(-) ^{***} (.1425)
<i># of Watchers</i>	-	-	-	(+) (.0002)	(-) (.0005)
<i>Average Views</i>	-	-	-	-	(+) (.0000)
<i>Content Length</i>	-	-	-	-	(+) ^{***} (.0000)
R-Squared	0.006	0.075	0.084	0.144	0.2475
# Obs	120	120	120	120	120

***p <.01, **p <.05, *p <.1

b) Talk Pages:

The analysis of talk pages further corroborates the previous findings. In particular attention should be paid to the change in Wikiprojects statistical significance when GA quality ratings are included as a dependent variable. While registered contributions and Wikiprojects are positive and highly significant in talk pages presumably belonging to top quality Wikipedia articles, the effect of Wikiprojects disappears at lower quality levels. In other words, the results suggest that Wikiprojects play an important role in facilitating coordination activity that impedes reductions in the quality of its top tier articles. However, Wikiprojects' lack of statistical significance in talk pages at lower quality levels along with the highly significant negative relationship between registered user reverts in content pages, and Wikiprojects statistical significance when coupled with anonymous contributions suggests that a lack of coordination in lower quality articles is resulting in registered users following their interpretation of what is deemed to be acceptable material.

Table 9:
Impacts of Registered Users on Talk Pages

Quality Dummy ≥ 4	1	2	3	4
<i>Log of Registered Edits</i>	(+) ^{***} (.0172)	(+) ^{***} (.0203)	(+) ^{***} (.0247)	(+) ^{***} (.0296)
<i># of Wikiprojects</i>	-	(+) (.0213)	(+) (.0201)	(+) (.0218)
<i>Log of Reg Deletions</i>	-	-	(-) (.0479)	(-) (.0591)
<i>Average Views</i>	-	-	-	(-) (.0000)
<i>Talk Length</i>	-	-	-	(-) (.0000)
R-Squared	0.227	0.236	0.254	0.255
# Obs	120	120	120	120

Table 10:
Impacts of Anonymous Users on Talk Pages

Quality Dummy ≥ 4	1	2	3
<i>Log of Anonymous Edits</i>	(+) ^{***} (.0274)	(+) (.0290)	(+) (.0433)
<i># of Wikiprojects</i>	-	(+) ^{***} (.0221)	(+) (.0248)
<i>Average Views</i>	-	-	(+) (.0000)
<i>Talk Length</i>	-	-	(+) (.0000)
R-Squared	0.061	0.106	0.121
# Obs	120	120	120

This could result for various reasons, such as the lack of properly defined stylistic guidelines or the issues of reverse causation resulting in a disinterest in lower quality articles. That is, Wikiprojects may choose to focus on articles which require little work to achieve the status of high quality, although the distribution of Wikiprojects throughout quality categories presented in Figure 1 would contradict this view. However, the evidence presented in that table should be taken with caution given the size of the sample employed and the inability to discern how many of the Wikiprojects in high quality articles were there before high quality was achieved. Alternatively, making the fairly realistic assumption that high quality pages receive a

larger quantity of views, the clustering of WikiProjects in high quality articles could also be interpreted as their acting to first eliminate the potential threats anonymous users present to high quality pages, and second to engage in the production activity that will increase lower quality ones. The answer could be determined by calculating the change in the quantity of WikiProjects for a given page caused by an increase in quality while controlling for page views. Unfortunately the answer to this remaining puzzle is precluded by a lack of available data.

Conclusions:

Previous research has emphasized the similarities between the design principles employed by long-enduring CPR institutions and Wikipedias institutional functioning and structure. This emphasis has resulted in attributing Wikipedias success to the application of these institutional design principles. Contrary to the case of common-pool resources where over consumption of resource units and the under provision of institutional arrangements that regulate individual behaviour result in suboptimal social outcomes, the main problem Wikipedia has had to overcome is adequately allocating its members efforts between content production and monitoring and exclusion activity. Efficient monitoring and exclusion allow Wikipedia to reap the benefits that come about from the application of open access institutional models in cases where the inclusion of larger quantities of individuals results in greater social benefits caused by a reduction in transaction costs. The results of this investigation partially corroborate this hypothesis.

The main finding of this paper is that WikiProjects play a significant role in high quality articles. This may be due to WikiProjects ability to facilitate the creation of stylistic guidelines. By doing so, they clarify what constitutes desirable editing activity thereby decreasing the costs of reaching consensus and excluding members that do not conform to their guidelines. The positive and statistically significant relationship between WikiProjects and high quality articles suggests that they play an important role in impeding reductions in quality caused by anonymous contributors. However, the allocation of too much effort to monitoring and exclusion activity in lower quality pages may decrease the potential benefits provided by anonymous users editing activity. The results show a negative impact of registered user deletions on an articles quality which suggests this may be the case. Thus the mere presence of an institutional context in Wikipedias case may be a necessary but insufficient condition for the provision

and creation of quality content. Thus a better understanding of the benefits conferred to Wikipedia by its institutional characteristics warrants further investigation.

References

- Alchian, A.A. (1989). *The invisible hand*. J. Eatwell, M. Milgate, P. Newman, (Ed.). New York, NY: Macmillan Press.
- Anthony D., Smith S.W., & Williamson T. (2009). Reputation and reliability in collective goods: The case of the online encyclopedia Wikipedia. *Rationality and Society*, 21(3), 283-306.
- Benkler, Y. (2001). Coases penguin, or, Linux and the nature of the firm. *The Yale Law Journal*, 112 (3).
- Butler B., Joyce E., & Pike B. (2008). Don't look now but we've created a bureaucracy: The nature and roles of policies and rules in Wikipedia. *Proceedings of the SIGCHI conference on human factors in computing systems*, Italy, FL.
- Forte A., Vanessa L., & Bruckman A. (2009). Decentralization in Wikipedia governance. *Journal of Management Information Systems*, 26(1), 49-72.
- Kittur, A. & Kraut, R. E. (2008). Harnessing the wisdom of crowds in Wikipedia: Quality through coordination. *Human-Computer Interaction Institute*. Paper 99.
- Lakhani K.R. & von Hippel E. (2003). How open source software works: "Free" user-to-user assistance. *Research Policy*, 32, 923-943.
- Layard P.R.G & Walters A.A. (1978). *Micro-economic theory*. United States: McGraw-Hill.
- Lerner, J. & Tirole, J. (2002). Some simple economics of open-source. *The Journal of Industrial Economics*, 50 (2), 197-234.
- Lerner, J., Pathak, P.A., & Tirole, J. (2006). The dynamics of open-source contributors. *The American Economic Review*, 96 (2), 114-118.
- Ostrom, E. (1997). A behavioural approach to the rational choice theory of collective action: Presidential address. *The American Political science Review*, 92 (1), 1-22.
- Ostrom, E. (2003). How types of goods and property rights jointly affect collective action. *Journal of Theoretical Politics*, 15, 239.

- Ostrom, E. (2008). *Governing the commons: The evolution of institutions for collective action* (22nd edition). New York: Cambridge University Press.
- Sandmo, A. (1989). *Allocation, information and markets*. J. Eatwell, M. Milgate, P. Newman, (Ed.). New York, NY: Macmillan Press.
- Viegas, B.F., Wattenberg, M., & Dave, K. (2004). Studying cooperation and conflict between authors with *history flow* visualizations. Retrieved from <http://www.research.ibm.com/visual/publications.html>.
- Viegas, B.F., Wattenberg, M., Kriss, J. van Ham, F. (2007). Talk before you type: Coordination in Wikipedia. Retrieved from <http://www.research.ibm.com/visual/publications.html>.
- von Hippel, E. & von Krogh, G. (2003). Open-source software and the "private-collective" innovation model: Issues for organization science. *Organization Science*, 14(2), 208-223.