THE PARADOX OF NOVICE CONTRIBUTIONS IN COLLECTIVE PRODUCTION: EVIDENCE FROM WIKIPEDIA

Andreea D. Gorbatai

gorbatai@haas.berkeley.edu

Haas School of Business

545 Student Services #1900

University of California at Berkeley

Berkeley, CA 94720-1900

Keywords: collective production, public goods, expertise

February 10, 2014

Electronic copy available at: http://ssrn.com/abstract=1949327

THE PARADOX OF NOVICE CONTRIBUTIONS IN COLLECTIVE PRODUCTION: EVIDENCE FROM WIKIPEDIA

ABSTRACT

This study uses the online encyclopedia Wikipedia to examine the links between expert producers of collective goods, demand for such goods, and good quality. Since collective production settings lack a price-like mechanism, producers do not have direct information about demand for goods so they may fail to produce goods that are needed. In this study I identify a social mechanism through which producers receive, and respond to information about consumer needs across a set of heterogeneous collectively produced goods. Using a longitudinal dataset of 187 million contributions to Wikipedia articles and article demand between October 2008 and February 2009, I model the contributions of novice and expert producers to article quality, and evaluate the relationship between consumer need and novice and expert contributions. Findings show that novice contributors have a direct negative effect on good quality, but their participation in producing a good motivates experts to contribute and increase the quality of the good, thus mediating the relationship between need for goods and expert contributions. These results provide evidence that collective goods fail to satisfy consumer needs in the absence of direct information from consumers, and highlight the paradoxical role of novices in providing a cue about these needs.

Electronic copy available at: http://ssrn.com/abstract=1949327

THE PARADOX OF NOVICE CONTRIBUTIONS IN COLLECTIVE PRODUCTION: EVIDENCE FROM WIKIPEDIA

Public goods are omnipresent in society, and they are inextricably linked to social order. Governments provide national security, clean air, and road infrastructure; communities— social norms and collective events (Coleman, 1990; Oliver and Marwell, 1992; Heckathorn, 1993; Kollock, 1998); and groups of organizations create new industries, communication infrastructures, and industry regulations (Fligstein, 2001; Monge and Contractor, 2003; Haveman, Rao and Paruchuri, 2007). This study focuses on an important type of public good provision— collective production— that entails coordinated action by a large number of independent agents, such as scientists engaged in collective production of a map of the human genome, musicians engaged in the production of a free music festival, or internet participants contributing to the creation of an online encyclopedia.

Over the past decade, modern technology has facilitated the creation of many online collective production platforms. A-synchronous global reach and technology-based coordination have helped interested parties find one another and coordinate their contributions to create large-scale successful platforms (Kollock, 1999). As of 2014, many online review sites, question-and-answer forums and support communities, link-sharing sites, open software production, and photo sharing and curating platforms host public goods that are available to the global audience of internet users (von Krogh, Spaeth and Lakhani, 2003; Lakhani and Wolf, 2005; Setia et al., 2012). Even in situations where the online platform creates a marketplace where one-on-one exchanges occur, such as the short-term labor market website elance.com, individuals contribute reviews that, in aggregate, offer information regarding the skill and trustworthiness of

participants, and indirectly, the overall marketplace. These online collective production settings provide a rich trove of digital data that allow researchers to explore questions regarding markets and categories (Kovács and Sharkey, 2013; Leung, 2014), gatekeeping and power dynamics (Shaw, 2012), project governance (Shah, 2006; O'Mahony and Ferraro, 2007), coordination and community structure (Dahlander and O'Mahony, 2010), and motivations to participate (Lakhani and Wolf, 2005).

Online collective production generates public goods that are available on a global scale. But even within the same collective production setting, not all goods are of equal interest to consumers. Some goods are needed more than others. Of particular concern here is the problem of misalignment between the need for collective goods, and producers' incentives to provide them. Since public goods are free and non-excludable, there is no price mechanism to align supply and demand. Instead, social rewards and intrinsic motivations drive the amount and allocation of effort in collective production endeavors. This means that goods that are important to consumers, such as particular open source software features or books of interest to a minority audience, may remain ignored by producers if the consumers' interests are not aligned with producers' motivations. Moreover, since collective goods are non-rival, the volume of goods is not a concern; multiple consumers can benefit from the same good without diminishing its utility to any of them. Given this, I suggest that a relevant metric to evaluate collective production success does not rely on the number or quality of goods produced but rather on the quality of goods that beneficiaries are interested in consuming.[1] One can imagine a situation where a free country music festival is organized, but the local audience does not enjoy country music, or a

[1] This refers to a non-mediated market where there are no intermediaries to influence and predict consumer tastes for particular goods. In the case of mediated markets, such intermediaries could arguably affect what collective goods consumers are interested in.

review website for local restaurants receives a lot of input on local coffee shops' quality while many visitors to the site are more interested in fine dining.

This study evaluates the conditions under which producers create high quality collective goods that meet the needs of consumers in absence of a pricing mechanism. I propose a social mechanism connecting producers and consumers of collective goods, such that consumers can cue producers as to what goods are most needed. In doing so, I draw on research on collective action and producer motivations to examine the actions and interactions of novice and expert collective good contributors. Based on the idea that expert contributors are partially driven by prosocial motives and that they can observe changes to collective goods, I propose and test a theory of collective good improvement by expert producers as a result of cues from novice producers being interpreted as consumer need. I empirically test the existence of this mechanism in the context of the online encyclopedia Wikipedia using a unique longitudinal dataset of 187 million contributions to article production coupled with data on article quality and demand for articles. My findings suggest that ignoring the relationship between producer and consumers of collective goods, and the role of occasional producers as intermediaries between these two parties, can result in negative consequences for a large and important category of public goods.

THE PRODUCTION OF COLLECTIVE GOODS

Economics and sociology provide a variety of perspectives on contributions to collective goods. While classical economic theory focused on the importance of selective incentives (Olson, 1965) and of material incentives for cooperation, recent research has also shown that social rewards such as social approval or prestige from participation affect incentives to participate in public

goods production (Soetevent, 2005; Ariely, Bracha and Meier, 2009). Additionally, recent work indicates that people contribute more to the public good than pure self-interested behavior can explain (Fischbacher and Gächter, 2008), and that some declines in cooperation can be attributed to "frustrated attempts at kindness" rather than deliberate free riding strategies (Andreoni, 1995). For an extensive review and a discussion of cooperation enforcement mechanisms, see Chaudhuri (2011).

Sociologists have proposed a wide range of antecedents for participation in collective good production, in settings ranging from social movements and volunteer work (Fernandez and McAdam, 1988; Musick and Wilson, 2007) to open source software and communities of practice (Lakhani and Wolf, 2005; Shah, 2006). By asking "Why do people participate in collective production?" research in these areas has yielded two main research streams. One stream has been largely concerned with the conditions under which people become involved in collective production while the other has examined motivations for continued participation.

Research concerned with initial contributions has argued that actors' initial mobilization in collective production is affected by cultural and structural factors (McCarthy and Zald, 1973; Snow, Zurcher and Ekland-Olson, 1980; Oliver, 1984; Gould, 1991; Gerhards and Rucht, 1992) and has shown that volunteer work, civic action and community involvement are performed by a non-representative set of participants from among beneficiaries of the goods produced (Heckathorn, 1993; Musick and Wilson, 2007). Research on public good contribution to electronic communities in organizations has yielded similar results (Borgatti and Cross, 2003).

Collective Production Participation and the Development of Expertise

Research on collective action aimed at political change or collective production of goods, such as open source software, also indicates that a few highly involved contributors develop expertise in the production processes that they are involved in, while many remain occasional, novice contributors or “token” participants (Hausknecht, 1962; McCarthy and Zald, 1973; Mockus, Fielding and Herbsleb, 2005; Anthony, Smith and Williamson, 2007). Those contributors who are highly involved have a higher propensity of cooperating toward the resolution of public good dilemmas and "share an orientation toward a collective or shared good that benefits others besides the participants" (Oliver and Marwell, 1992). Collective actions organized around markedly different goals and tactics are often mobilized in markedly similar ways.

In the context of open source software production, researchers have similarly found that only a small percentage of contributors are highly involved and experienced with the collective production process (Shah, 2006; Kriplean, Beschastnikh and McDonald, 2008; Panciera, Halfaker and Terveen, 2009). These contributors have extensive expertise in the process of collective production, are often involved in high-level coordination and integration activities and are experts in the collective production process (Dahlander and O’Mahony, 2010). They report participating because they derive enjoyment from their participation, and enjoy feedback from their peers. These expert contributors are receptive to bug reports, suggestions and questions from product users, due to altruistic motives—knowing that they helped create a product that is useful to others. Conversely, a large number of peripheral contributors participate only occasionally or briefly in collective production, helping with a particular product feature or with product distribution (Kriplean, Beschastnikh and McDonald, 2008).

Research on communities of practice has identified a temporary position of "legitimate peripheral participants" for less involved contributors, and proposed that this position is overcome as these contributors socialize and learn to become more effective members of the community (Brown and Duguid, 1991). Studies suggest that highly involved public good contributors and peripheral ones have different motivations and different participation "careers" (Kriplean, Beschastnikh and McDonald, 2008; Panciera, Halfaker and Terveen, 2009). Inexperienced participants make occasional, localized contributions, are novices from the point of view of understanding the coordination work that enables successful collective production, and are not interested in gaining expertise (Shah, 2006). Conversely, highly-involved, expert contributors have deep knowledge of the production process and are qualified to further the goals of the collective production endeavor (Dahlander and O'Mahony, 2010). Therefore, when both expert and novice contributors are free to alter the quality of the collective good, I predict:

Hypothesis 1. On average, expert (novice) contributions increase (decrease) the quality of collective goods.

Motivations and Collective Production Participation Patterns

The literature on collective production has extensively examined the conditions under which individuals become involved in collective production, and the relationship between individual motivation and contribution patterns. Relatedly, this research has stressed the importance of selective incentives for continued participation in public good provision (Olson, 1965). Among such selective incentives, identity, prestige and other social rewards have received extensive attention in the sociological literature (Gould, 1991; Whitmeyer, 2007; Willer, 2009).

A major source of social rewards is one's peer group. The act of participating itself "creates new social ties, even as it relies on pre-existing ties as a source of solidarity" and initial mobilization (Gould, 1991). In absence of social ties to co-participants, continued commitment is difficult to sustain, because participants may lack peer rewards and a sense of collective identity (Gould, 1991). The new social ties that arise as a by-product of collective production participation can generate a series of competitive and non-competitive social rewards (Goode, 1979; Coleman, 1988; 1990; Willer, 2009) and selective incentives bestowed on contributors (Olson, 1965; Hardin, 1982; Grant and Gino, 2010) and motivate further contributions (Howison et al., 2011). In addition to social rewards from participation, contributors may also be intrinsically motivated through a series of psychological factors such as self-efficacy, enjoyment, learning, challenge or escalating commitment (von Krogh, Spaeth and Lakhani, 2003; Shah, 2006). Personal valuation of an end-product for its impact on society—such as political change, anti-discrimination laws, clean air, safe neighborhoods, transferable training or improved laws and regulations—are other motivations shared by many participants in the collective good production process (Traxler, 1993; Oberschall, 1994; Johansen, 2002; Lakhani and Wolf, 2005).

While studies of collective production are informative regarding participant careers and motivations, fewer asked: "Are the public goods produced useful to their potential consumers? Do they meet consumer need?" In order to address this question, I invoke a less-used classification of public goods: the distinction between homogeneous goods that everyone values to some degree, and heterogeneous goods, for which potential interest in production and consumption cannot be placed on a continuum (Hardin, 1982). For example, goods such as open source software, Congressional bills, education or health care reform are heterogeneous public goods where individual interest ranges in terms of the desired attributes of the public goods.

Image processing open source software may be a priority to some consumers, while others may be uninterested in image processing but want a powerful video editing open source software program. Even among those interested in image processing software, some consumers may be interested in prepackaged processing options while others may want flexibility in making editing decisions. Research often theorizes public goods as homogeneous and focuses on the extent to which free rider dilemmas are overcome and production is successful; however, collective action research indicates that individuals are often heterogeneous both with respect to the resources they have available for collective good production, and with respect to their interests in the various attributes of the produced goods (Olson, 1965; Marwell, Oliver and Prahl, 1988; Heckathorn, 1993). In other words, many of the public goods we study have heterogeneous characteristics.

At the same time, expert contributors are not a random sample from among potential participants: they have the resources to contribute and a set of motivations different in kind from the motivation of non-contributors. Additionally, they are likely to receive social rewards from their peers (Coleman 1990), which may skew their contributions away from broader consumer needs. Given that collective production experts have the know-how to produce public goods, but that their incentives are not aligned with the beneficiaries of their production efforts, the key question follows: “Under what conditions do expert contributors produce useful goods?” Building on the literature regarding motivation in collective production, which suggests that sometimes individuals participate in collective production as prosocial activity due to altruistic motivations to satisfy the needs or interests of others (Grant and Gino, 2010), I suggest that when expert producers observe or infer consumer interest in a particular characteristic of a collective good, they are more likely to expend effort towards producing the desired characteristic:

Hypothesis 2: All else being constant, expert producers are more likely to contribute to public goods that consumers indicate are needed.

Several mechanisms can serve to align production with the need for public good production. Most notably, in political science and public choice economics, voting mechanisms play a role in aggregating constituent demand (Holcombe, 1989). Similarly, election-based selection of committee members can ensure that the interests of the beneficiaries are adequately represented (McElroy, 2006). In virtual communities, aggregating reviews, comments, questions, or other feedback from consumers can provide a measure of interest in information-type public goods (Sun and Zhu, 2012), and being a high-profile producer of goods that address these needs can lead to selective incentives in the form of status in the online community (Lampel and Bhalla, 2007). More importantly, the need for a particular collective good or attribute is not necessarily expressed as a request. In knowledge-production settings, need may be inferred by producers from consumer behavior, such as downloads of papers, citation counts, comments, ranking of most popular articles, or re-posting content on other forums. In software development, developers can report bugs and discuss problems on discussion boards, highlighting for expert producers those areas where the software needs additional development (Setia et al., 2012).

I propose that producers use signs (Gambetta, 2009) of consumer interest or need to modify their beliefs about the utility of a particular public good. For example, radio station hosts or blog writers may alter the content of the knowledge produced in response to signs of interest in a particular topic (Sun and Zhu, 2012). Because such producers often lack direct knowledge of latent demand, such as the number of individuals interested in consuming content on a particular topic, they may approximate this demand and redistribute resources over a set of topics using cues from the audience, such as listener calls, posted comments, or ratings. In situations where

collective producers are willing and able to produce a range of collective goods or collective good attributes but do not have information to help them prioritize the resource allocation, a mechanism that provides information about beneficiary needs may increase producer contributions in that direction. While "we cannot take for granted that signs [of interest] are noticed" (Gambetta, 2009) I predict that, when available, signs of interest from potential beneficiaries are used by producers to adjust effort allocation, such that experts respond to signs from novice contributors instead of consumer need.

Hypothesis 3: Novice contributions mediate the relationship between demand for a good and expert contributions.

RESEARCH SETTING: THE ENGLISH WIKIPEDIA

To test my predictions I examine article production patterns in Wikipedia, a free online volunteer-contributed encyclopedia and a salient example of collective production. Three main considerations are behind my choice of research site. First, as described below, English Wikipedia represents an ideal case of collective good production because as of 2013 volunteer contributors have created over four million encyclopedia articles. This offers an opportunity to compare and contrast contributions to the collective production process across heterogeneous but similar goods. Second, by the nature of its online platform, Wikipedia collects an unprecedented amount of longitudinal data on both actions and interactions of contributors, and on views of encyclopedic articles (consumer need), while expert producers are unable to directly observe

article views.[2] Third, encyclopedic articles are a type of good that requires minimal skills to consume (i.e., literacy and online access), such that one can assume an individual who views a page has "consumed" the existing information.[3] Taken together, all these characteristics recommend Wikipedia as an ideal research setting for the proposed collective production theory. In the following section, I elaborate on contribution patterns and processes in the English Wikipedia.

Wikipedia participation process. The success of Wikipedia relies on a technology called wiki software. Individuals can modify any existing page for everyone else to see, while previous versions of the page remain accessible via a "history" page.[4] Individual contributions are not censored or screened before being included in the encyclopedia, such that anyone can contribute.[5] For this reason, Wikipedia has attracted over six million registered contributors who produced over 4 million articles in English and over 19 million articles total in more than 270 languages by March 2013.[6] As of 2009, Wikipedia was also registering approximately 477 million views per day, half of which were to English-language pages, making it the seventh most visited website in the world.[7] However, many consumers of Wikipedia's content (readers) remain unaware that they can contribute. Of those who create accounts, most do not make more

[2] As of May 2010, Wikipedia article history pages started to include a link that allows anyone to access the page requests for that article over time. While this time period falls outside my data, and I have no data on the use of this information by expert contributors, qualitative data suggests that experts are unlikely to seek it.

[3] For many public goods there may be individuals who do not benefit from them unless they have the ability to consume them; for example, individuals cannot benefit from engineering specifications or chemical formulae unless they have the ability to use them, and the necessary tools (Drahos, 2004). Since Wikipedia exists online, individuals without online access would not be able to consume this good, and some individuals consuming Wikipedia articles may be unable to assess their quality, or may be satisfied with a lower good quality than others – in the same manner as asthmatics need a higher air quality than healthy individuals do. This study is not concerned with issues such as online literacy, or digital divide or inequality in access to the internet and in internet skills (Schradie, 2011).

[4] During the period spanned by my dataset, January 2001 to May 2009.

[5] During the period spanned by my dataset, January 2001 to May 2009.

[6] http://en.wikipedia.org/wiki/Wikipedia:Size_comparisons. Retrieved on March 1, 2013.

[7] Data retrieved on October 25, 2010 from Alexa Traffic Rank 2010. http://www.alexa.com/topsites and http://stats.wikimedia.org/EN/TablesPageViewsMonthly.htm

than two or three contributions (ever) to articles; less than one in five registered contributors contributes more than ten times.[8] This is similar to open source software, where the expert (core) developers produce 88 percent of the code (Mockus, Fielding and Herbsleb, 2000).

In order to understand the dynamics of collective production, I first engaged in participant observation of contribution to Wikipedia by performing Wikipedia edits and observing other contributors' edits, and discussion threads regarding both general policies and coordination in article writing between December 2006 and May 2011. During this time I also interviewed a random sample of 35 experienced (expert) contributors to Wikipedia.[9] By analyzing the interview data, I have singled out qualitative evidence pertaining to the production processes and the article-based interactions among contributors.

Wikipedia structure and rules. Large-scale reciprocal interdependence requires a large coordination effort. The increase in the number of articles in the English Wikipedia from 100,000 by the end of 2003 to over 3 million by 2009, coupled with the increase in the number of registered editors to 1,824,439 as of December 2007, led to the proliferation and increased complexity of Wikipedia's structure and contribution policies and norms. Even ostensibly simple policies such as "Wikipedia is not a place for original research" or "Always strive for a neutral point of view" have been subject to debate and increasingly refined or expanded in scope (Butler, Joyce and Pike, 2008). As a result of these coordination challenges, fewer than 35% of Wikipedia pages are dedicated to article content, while the rest represent discussion, policy, and

[8] More than 80% contributed less than 10 times, and only about 3% edited articles more than 100 times. About 1% of registered contributors made more than 500 edits to articles.

[9] To select my interviewees I started with a theoretical sample of 50 registered contributors who had contributed between one and 100 times to article writing (novices), and 94 who had participated over 100 times (experts). All of the novices either failed to reply or politely declined to be interviewed; approximately one-third of the experts were successfully interviewed. Interviews were semi-structured, based on an interview guide which touched upon the participants' first contributions, their current contribution practices, and, if applicable, their departure from Wikipedia. Interviews were conducted live via VoIP and IM, with the exception of three contributors who preferred e-mail.

user profile pages.[10] This is consistent with studies of other studies of collaborative user-generated content systems: while the likelihood of a high quality result may increase with the number of participants, costs associated with cognitive and coordination costs increase even more due to information overload (Ransbotham, Kane and Lurie, 2012) resulting in more policies and rules regarding collaboration.

Wikipedia and novice contributors. According to my interviewees, the majority of Wikipedia contributors start by chance, becoming involved in collective production by way of articles they have consumed.[11] This information suggests that the more demand for an article, the more likely it is the article will receive contributions from novice producers who can make a low-cost contribution at the time of reading. The low-cost contribution assumption is supported by my interviewees, who stated that first contributions were either content that involved minimal effort, such as adding or correcting information about one's favorite band, native town, or alma mater (ten interviewees), or minor copy edits (ten interviewees).

Consistent with other collective production participation theories, many interviewees stated that initially they did not know how to communicate with others, contribute useful work, or even retrieve their own contributions.[12] Overall, as novice editors they were not aware of the collaborative process through which article writing took place. They could not decipher the history of articles by examining auxiliary pages, and they were not cognizant of the many ways

[10] The English Wikipedia had 11,405,052 pages by the time it had 2,183,496 articles. Less than a fifth of pages are articles because *Discussion*, *Editor* and rule pages are included in the page count.

[11] Informal discussions with other Wikipedia contributors, both experts and novices, suggest that most of the initial contributions are generated as a by-product of consuming (reading) articles. Research on other forms of collective action suggests that individuals benefiting from collective action may similarly lend a hand if they are accidentally exposed to the opportunity to produce, or recruited by more active participants.

[12] Some novice contributors are anonymous, while others may have registered and use a username. While registered contributors can easily retrieve their past contributions based on their unique username, anonymous contributors can rarely do so, because the IP addresses used automatically as substitutes for their usernames are often impermanent (dynamically allocated to internet users). Even when the possibility to retrieve one's contributions exists, lack of interest or skill may still preempt novice contributors from receiving feedback and learning from their work.

in which they could contribute to Wikipedia or of the rules and policies governing these contributions. Evidence from Wikipedia indicates that many of the novice editors who register contribute only a few times (other novices contribute anonymously), and never become engaged in the collaboration process. Given that repeated contributions lead to learning and socialization, I assume a positive relationship between the number of contributions to Wikipedia articles and expertise in the production and collaboration process.

Experts' motivations, collaboration processes, and article quality. Wikipedia remains in principle an encyclopedia that anyone can edit, but its increasing complexity means that becoming an expert contributor requires an increasing amount of effort and dedication to understanding the rules by which the community functions and the types of legitimate and appreciated work (Kriplean, Beschastnikh and McDonald, 2008). This is similar to other forms of public good production, such as basic science or open source software: while "anyone" can contribute, experts are aware and capable of using institutionalized processes - such as submitting research to peer-review journals or committing code through the appropriate channels - to improve the quality of public goods. In Wikipedia, features such as the fragmentation of the same discussion across multiple pages, the use of notice boards located in hard-to-find locations, and intricate user policy systems create "private spaces for [expert producers] to act away from the eyes" of novice producers. This has the effect of keeping the latter away in the same manner that the "law in action" makes it difficult for ordinary citizens to execute their rights despite the fact that "the law in the books" is publicly available (Oz, 2009). Wikipedia researchers have argued that such spaces enable expert contributors to selectively engage in certain discussions and help increase the speed and efficiency of information exchange among experts at the expense of broader, novice participation (Oz, 2009).

Wikipedia "experts" are therefore not necessarily knowledge experts but individuals with experience in the article production process, who understand the contribution process and are privy to and often involved in Wikipedia's "private sphere," similar to the manner in which experienced decision committee members in universities or neighborhood activists may be aware of information that is publicly available but not readily accessible to infrequent participants and non-contributors. In contrast, for most consumers and novice editors—who do not know about Wikipedia's social processes of collaboration—Wikipedia is not a community or a site of production but rather a source of information. Many forms of collective goods production are similarly structured around a core, or a set of foci of intense activity, with consumers who "never initiate action, but only [occasionally] respond to the opportunities created by [expert participants]. [Moreover], it is not certain that they will contribute, even if they are asked" (Oliver and Marwell, 1992).

As is the case with other collective production settings, expert contributors are fueled by enjoyment, personal growth or need, and desire to help others (Nov, 2007; Yang and Lai, 2010). In addition to intrinsic rewards, expert contributors receive selective incentives from peers in the form of status, formal administrative positions and recognition, either as electronic messages or as publicly displayed "barnstar" rewards[13] (Ciffolilli, 2003; Kriplean, Beschastnikh and McDonald, 2008; Restivo and van de Rijt, 2012). However such peer recognition can occur only after one has gained experience as a member of the community. My interviewees explained that as they contributed frequently to article writing, they became aware of the presence of other contributors, and commenced learning about the channels and norms to communicate with them

[13] Barnstars are a type of merit badges that contributors award one another for extraordinary work. They are publicly displayed on the receiver's personal page. Research has shown that barnstars are effective as motivators for future participation and that they are recognized as a source of status among expert contributors (Restivo and van de Rijt 2012).

and to receive feedback on their work. Thus expert contributors participate in writing articles they are interested in learning more about, articles that other experts ask them to consult on, and articles that they consider important.

The ability to locate other experts and contribute, communicate and collaborate in article production according to the norms and rules espoused by the community enables expert producers to improve the quality of articles. In contrast, as novice contributors make article changes they are largely unaware regarding the process of production of the articles and others' expertise levels. They are unfamiliar with the rules and policies regarding article writing and with the social norms governing contributions to articles, and unaware or uninterested in locating information about these policies, rules and norms. They are also unable to request assistance as they lack social capital in the community and sometimes even the ability to access the communication channels the experts employ. For these reasons, novice contributions may often harm articles: they may be redundant with information in another section, lack proper references, be biased in their wording, or simply be irrelevant. Often novices make small contributions such as adding a comma or capitalizing a word; even at this level of detail, they may detract from overall article quality because they are unfamiliar with Wikipedia's Manual of Style and make contributions that do not follow its rules.

Experts see novice participation as a sign of article interest. Wikipedia expert contributors have two means of observing changes to articles within their range of interest. First, they can periodically revisit any article and examine the history page, which contains an index of all modifications to a specific article. Second, to facilitate collaboration in production, Wikipedia offers contributors the option to monitor changes to articles they are interested in through a

dashboard page called the "watchlist." Many interviewees reported actively monitoring hundreds or even thousands of articles,[14] and frequently visiting articles modified by novices.

The interviewees clearly stated that novice contributions to articles serve as an indicator that readers were interested enough to have read the article attentively and found ways to improve it: "I'm a fast writer, so often miss punctuation or spelling errors, which other people will fix... It's neat seeing that people are reading [what I write so carefully], and I can tell it's not someone using an automatic checker for typos. It is great knowing people are reading it and paying close enough attention to [see] typos." Another interviewee explained: "Some kinds of [novice] edits are good in themselves but do not conform to the style or to the coherence of the article. [Regardless, things] like fixing my typo shows someone is reading and paying attention—that is motivating. [Novices] will write … 'This aspect of the article needs more coverage' and sometimes they are completely right and that makes me … add more content."

My interviews suggest that although novice contributions to collective action may consist of small, inconsequential changes, their participation signals to expert producers that certain goods are of interest to consumers. These contributions are analogous to radio listeners calling in to a particular show, or inexperienced stakeholders anonymously suggesting changes to a particular organizational policy being debated. In order to empirically test the mechanism proposed, I proceed with describing the quantitative dataset, my methods, and the findings of my analysis.

[14] As the interview data suggest, experts are monitoring articles for changes with the primary purpose of preserving the quality of the article. However, if experts respond to a substantive change in the article by better integrating it in the original text, or by improving the writing, they could incidentally improve the quality of the article as well.

METHODS

The comprehensive panel dataset employed in this study was created through the merger of several unique data streams provided by volunteer Wikipedia contributors at the author's request and public data made available by the Wikimedia Foundation. The final article-interval level dataset was created using five separate data streams which include (1) the complete history of over 185 million contributions to over 3.5 million English Wikipedia articles between January 2001 and May 2009, (2) a record of 2,592 hourly intervals of all Wikipedia article requests received by Wikimedia servers between October 1, 2008 and January 31, 2009, (3) a dataset indicating the number of contributors monitoring each article as of October 2009, (4) article length and quality ratings as of May 2010, and (5) knowledge categories for each article as of October 2010.[15] For computational reasons, my analyses use a one percent random sample of articles from this dataset, which contains 168,739 article-interval records for 21,986 articles,[16] where an interval represents a half-month period for the production (article edits) and consumption (article views) data between October 2008 and January 2009.

Expert contributors on Wikipedia are those who have experience participating in article writing. Given the nature of article writing in Wikipedia, which consists of synthesizing information from published materials, these contributors are not necessarily content experts, but process experts who possess knowledge of Wikipedia policies and norms regarding contributing and collaborating. While no absolute cutoff point exists between novice and expert contributors, I have chosen 100 contributions as a cutoff point to separate expert from novice contributors,

[15] Due to data shortcomings, only 3,712,980 (20%) of the total article-interval records contained data on demand (article views), and 1,273,143 (53%) of articles had cross-sectional information about monitoring patterns. Missing data problems arose randomly with respect to the mechanism of interest, from formatting problems with article titles and data collection issues (server failures).

[16] About 90% of these articles were started before October 1, 2008, such that the data includes all eight intervals; the remainder has fewer than eight time intervals of data.

regardless of whether novices have a registered account or contribute anonymously; this 100-contribution threshold is consistent with other research on expert contributor to Wikipedia (Howison et al., 2011).

Given that my dataset contains longitudinal data on article production and consumption, and cross-sectional data for article quality, I employ a cross-sectional analysis of article quality for the first hypotheses, followed by a longitudinal analysis to examine the relationship between expert and novice contributions.

Hypothesis 1. I test Hypothesis 1 using a logistic regression model to evaluate the extent to which expert and novice contributions, log-transformed, predict the quality of the good produced, and the extent to which consumer need for goods fails to align with the quality of goods produced. On Wikipedia the categories of article quality are, in decreasing order: Featured (exemplary) article (FA), A-class article, Good article (GA), B-class article, C-class article, Start, and Stub, where Start articles are usually only about one paragraph long, and Stub articles contain at most a few sentences. Article assessment for factual completeness takes place after an article is classified as belonging to a WikiProject,[17] in which a set of participants interested in a broader subject related to the article's topic evaluates existing articles on that topic and coordinates plans to improve them.

Although Wikipedia employs a 1-7 scale to evaluate article quality, I use the definition of this scale to create a binary variable to reflect the extent to which an article is likely to satisfy consumer need, where articles with quality of one meet a minimum requirement (B-class or

[17] According to Wikipedia, "a WikiProject is a project to manage a specific topic or family of topics within Wikipedia. It is composed of a collection of pages and a group of editors who use those pages to collaborate on encyclopedic work." WikiProjects help coordinate and organize the writing of those articles. More than half of Wikipedia articles were rated for quality by at least one WikiProject.

more) that "readers are not left wanting, although the content may not be complete enough to satisfy a serious student or researcher" (Wikipedia 2011), and articles graded zero fall short of this criterion.[18] In addition to article quality ratings, I use a cross-sectional negative binomial model to test the relationship between article length at the end of the last interval and cumulative expert and novice contributions, log-transformed. Length measured as the number of characters represents a reasonable metric of the volume of information a consumer receives on a particular subject, although one cannot evaluate the comprehensiveness of the information as well as article structure, bias, or proper referencing based on length alone. By comparing the results across the quality and length dependent variables, I can then assess whether both experts and novices contribute on average by adding information to the article, thus increasing article length, and whether experts increase article quality while novices have a direct negative effect on it.

Hypotheses 2 and 3. In order to test Hypotheses 2 and 3, I model the number of edits by expert editors to article k during time interval (t+1) as a function of interest in the article and novice contributor editing during time interval t, and time-variant and invariant characteristics of the article. Here my dependent variable — expert contributions — is a count variable taking only non-negative integer values. Since linear regression models assume heteroskedastic, normally distributed errors, and these assumptions are violated when using count data, I employ a Poisson regression approach (Hausman, Hall and Griliches, 1984). The variance of the expert contribution variable is much greater than its mean, which is indicative of overdispersion, so I assume a negative binomial distribution.

[18] Article quality standards are clearly defined, both in terms of objective criteria and subjective reader experience. Quality is evaluated internally by Wikipedia experts according to community standards, and it is open to contestation by anyone concerned. Kittur and Kraut (2008) tested and confirmed the external validity of quality evaluations using ratings by non-Wikipedia participant readers.

I estimate within-group, fixed-effects negative binomial regression models, such that variation across articles is controlled for. This considers only within-article variance in the estimation of regression coefficients, so that the measured effect of consumption is independent of any time-invariant unobserved attributes of the article. The choice of testing Hypothesis 2 by estimating expert contributions in the period following novice contributions is based on the rationale that contributions closely following those of novices may simply erase the latter's work without stimulating additional contributions. Given that the majority of erased edits on Wikipedia occur soon after a contribution is made, the half-month interval seems sufficient to capture the long-term effects of novice participation on expert contributions. In the next section I explain my independent variable definitions, and then examine the results of the estimations, followed by a discussion of robustness checks and implications of my research.

Independent variables. Having described the models and the operationalization of the dependent variables, I now turn to describing the independent variables. Revealed demand for a collective good, measured here as *consumption* (reading) of article pages, is a key variable in this study. Although we do not have a direct measure of demand for Wikipedia articles, I argue that article views provide a good estimate of this because (1) articles are mainly text, such that most online visitors who locate them should be able to "consume" them; (2) articles are free; and (3) about 90% of Google search engine queries returned a Wikipedia article as a top link, and about 96% of searches returned a Wikipedia article in Top 10 (first page) results as of late 2008, which means that demand for a particular knowledge topic coming from internet users is likely to be reflected in Wikipedia page views.[19] Because the distribution of views is highly skewed, with a few widely-read articles and many more that are rarely read, this independent variable was log-

[19] As of 2008, more than 60% of article readers arrived at an article from a search engine, and the rest from links in other Wikipedia articles, Wikipedia's internal search engine, or links in other texts.

transformed. The other two important variables, *the number of novice edits* and *expert edits*, were log-transformed for the same reasons when used as regressors.

Given that we would expect the number of contributions, as well as article quality and length, to vary with the current state of an article, I include several control variables to account for the past history of the article in terms of the number and types of previous contributions. In some models, controls are employed for the cumulative number of previous edits (*cumul.edits*), while in others I control separately for edits by expert editors (*cumul.experts*), and edits by novice editors (*cumul.novices*). When articles are not protected, they are at risk of unintended damage or outright vandalism by other editors. The *protected* variable accounts for the extent to which the article has been protected in response to malevolent attempts to damage it, especially those coming from anonymous contributors.[20] When an article is protected, it cannot be modified by anonymous contributors or by editors with accounts created in the previous two days. Removing damage from an article and restoring the article to its previous state is called *undo*. A control for undo'ing was included in some models; undo edits, which are simply erasures, contribute less to article quality and length than other edits. The number of *experts* participating during a time interval was also considered as a control when estimating the number of expert contributions in the subsequent period, based on the assumption that participation by multiple experts may generate additional expert contributions as a result of iterative work.

Monitoring is an important variable for understanding expert editing patterns. Any registered contributor may monitor an article, which means that one is automatically informed when that article has been altered. Therefore, the more people monitoring an article, the more

[20] Ideally one would control for the period of time that an article was protected, but data limitations only make it possible to know whether or not an article had been protected without information about the duration of protection. Given that my definition of novices includes registered editors with fewer than 100 edits, semi-protection of an article would not have precluded all novices from making contributions to it.

likely it is that someone will react to a new contribution by making edits.[21] The *age* of an article is measured as the log-transformed number of days between the creation of the article and the last date in the dataset. Another important set of article attributes that I control for are labels assigned by editors to articles, such as *categories*. My data aggregates such category information to one of 24 high-level categories such as Business, Science, History, or Geography. A related control is *projects*, which measures the number of projects that a page is part of; for example, a page like "Albert Einstein" is part of both the WikiProject Germany and the WikiProject History of Science, among several other projects. Membership in multiple projects could be a confounding factor in the analyses because an article that touches upon multiple knowledge areas may result in more demand, elicit more contributions, and eventually lead to a longer article. *Importance* is a WikiProject rating reflecting the extent to which the article is considered central to that topic. It ranges from 1 to 4, where, by definition, top importance ("4") articles are a "must-have" for an encyclopedia, while high importance ("3") articles contribute information central to a knowledge area.[22] Articles labeled as important may attract more contributions from participants. In addition, very important articles may be of interest to more readers, such that one would expect a higher number of first-time edits to them. For this reason, I created a control for *first-time* edits, or the number of edits coming from participants who are contributing to Wikipedia for the first time since registering their username.

To account for variation in the distribution of work by editors on articles, I created *editor50%* to represent the number of contributors to an article ranked by their total edits to the

[21] Unfortunately, Wikipedia does not make public the names of editors monitoring each article, so I cannot distinguish between expert and registered novice editors who monitor an article.

[22] In order to preserve the same number of observations across models, whenever a variable such as quality, length, views, importance, monitors, ratio no comments, or ratio minor had missing values or was undefined I create a dummy variable to control for these cases. The controls do not affect the final results; the model estimates yield the same results without controls, and the estimates are available upon request.

article such that the sum of their edits is at least half of the total contributions by editors to that article. For example, if out of 100 edits on an article 20 edits come from editor A, 20 from editor B, and 17 from editor C, then *editor50%* would be 3. This variable indicates the extent to which the article was created through extensive peer collaboration versus a production process spearheaded by one or two individuals.

RESULTS

In brief, the proposed mechanism states that, on average, novice contributors to public good production negatively affect good quality while experts' contributions improve quality of goods but not necessarily of the most needed ones; paradoxically, novice contributions communicate to experts that the goods the latter have contributed are of interest, thus indirectly generating further good improvements. The results below provide strong support for these hypotheses and hence for the social mechanism improving alignment between customer needs and public goods. The positive correlation between Wikipedia article views and quality (0.22) and, respectively, views and article length (0.35), indicates that, given consumer participation and expert response, the quality and volume of goods in demand are higher than those of goods that are less demanded.

Insert Table 1 about here

Effects of novice and expert production on article quality. Table 1, models 1 through 3, presents the results obtained from the logistic regression estimates of the effect of novice and

expert contributions and demand for the article on article quality.[23] The results confirm that expert contributions have a statistically significant and positive effect on article quality, while novice contributions have a statistically significant and negative effect on article quality. The addition of control variables in models 2 and 3 slightly reduces the positive effect of expert contributions and the negative effect of novice contributions; however, the coefficients of interest remain statistically significant across all models. Hence, an increase in the number of expert contributions to an article increases the likelihood that the article is high-quality, whereas an increase in the number of novice contributions decreases this likelihood. This suggests a tradeoff between the fact that collective production needs consumers to participate in production to signal interest in a certain good, and the fact that too much novice (consumer) participation may decrease the quality of the goods.

Table 1 models 4 through 6 reports the results obtained from negative binomial estimates of the effect that novice and expert contributions have on article length. These models suggest that both expert and novice contributions increase article length. There is, however, a difference in magnitude and the disjunctive confidence intervals of the two coefficients: one additional unit in log-transformed expert edits corresponds to a 47-50% increase in article length compared to a 14-20% increase in article length for one unit in log-transformed novice edits. This suggests that expert contributions have a significantly stronger impact on article length than novice contributions, possibly because experts often contribute additional, substantial information, whereas many novices often make minor contributions.[24] Overall, results in Table 1 support the

[23] Using an ordinal logit estimator and the raw values of quality rankings, from Featured Article (quality = 7) to short and substantially lacking Stub articles (quality = 1), I obtain results that are consistent with the presented findings. These results are available upon request.

[24] A Poisson QML estimator was used to test the robustness of these findings; the results are statistically significant and strongly support the hypotheses. These models indicate a stronger effect of expert contributions and a weaker effect of novice contributions on article length. They are available upon request.

theory that while both experts and novices make positive contributions to article length, only expert contributions have a positive impact on article quality, while novice contributions decrease the quality of the article.

While the finding that experts increase article quality is hardly surprising, one could imagine situations where experts had already made all possible contributions to the collective goods. In this situation, novice contributions stimulate an expert response but do not result in an observable improvement of the collective good. Another situation would be a case where the production of the collective good elicits both strong positive and negative externalities, such as, for example, controversial legislation like capital punishment, or controversial organizational policies that differentially affect social categories of employees. One can imagine that experts' advocating for either side, receiving input from novice producers regarding these issues may be unable to satisfy the need for these goods because of opposing interests by other experts.

Insert Table 2 about here

Effects of demand and novice production on expert production. I show tests for Hypotheses 2 and 3 in Table 2 using four different regression models with fixed- and random-effect negative binomial estimators. The dependent variable in Table 2 is the number of expert editors' contributions to article k during time (t+1).[25] All covariates are lagged by one time

[25] In this study I infer an expert response to novice contributors by examining changes in experts' editing patterns on articles following novices' contributions; one may question whether the connection between the two has been misrepresented. To address this shortcoming, I examined expert editing in the two-week period following novice contributions, rather than during the same period. If the relationship between the two types of edits did not exist, or if it was restricted to erasing novice edits, effects two weeks afterwards would not be observed. The existence of the

period to control for the fact that during the same time period when novices made edits, expert editors may have responded by rejecting them, with no further contribution to article development.[26] In models 7 and 8, only time-variant controls were employed, due to the use of fixed-effects regression estimators. Models 9 and 10 include time-invariant controls such as the number of projects the article belongs to, article importance, and the number of monitors; their interaction with article demand; and article category.

The results in Table 2 confirm that Hypothesis 2 is strongly supported across models 8 to 10. Edits by novice contributors have a statistically significant and positive effect on contributions by expert editors, ranging from about 30% in fixed-effects negative binomial models to approximately 4.5% in random-effects models. Hypothesis 3 is strongly supported in models 7 and 8 which suggest that the positive direct effect of article consumption on expert editing patterns is fully mediated by novice contributions. Results presented in this table support the theory that experts are unaware of demand but they are stimulated to respond to article consumption if consumers signal demand for that particular good through their contributions as novice producers. While the current dataset contains no information that can help further disentangle expert motivation to respond to novice contributions, other Wikipedia studies suggest that expert contributors are driven mainly by prosocial motivations / ideology of sharing, and enjoyment / fun and social rewards (Nov, 2007; Schroer and Hertel, 2007; Yang and Lai, 2010). Research has found that expressions of gratitude for expert contributions result in an

---

effect in my quantitative analyses together with the reports of experts' reactions to novice contributions in my qualitative data suggests that the response to novice editing persists in the long run.

[26] Work by the Wikimedia Foundation's Erik Zachte has documented that more than one in four edits contributed by anonymous editors to English Wikipedia articles are erased, often immediately after they occur. Retrieved on October 27, 2010 from en.wikipedia.org/wiki/File:Erik_Zachte,_Edit_and_Revert_Trends,_Wikimania_2010.pdf

increase in subsequent contributions by as much as 60 percent (Restivo and van de Rijt, 2012) which suggests that social rewards are an important driver of participation.

Demand and novice contributions. The social mechanism proposed relies on the assumption that novice and expert motivations to contribute are different, and that novices contribute to article-type public goods that they are interested in consuming. In order to examine the latter assumption I used fixed-effect negative binomial estimators to evaluate the relationship between change in demand for Wikipedia articles and change in the likelihood of novice contributions within the same time period. The analysis confirms a positive correlation between the two, which is consistent with my qualitative research findings suggesting that novices contribute to articles mainly as a consequence of consuming them. These results are available upon request.

DISCUSSION AND CONCLUSION

This study contributes to our understanding of collective production by highlighting the micro-level mechanisms through which useful, high quality goods are produced. In particular, in addition to considering whether coordination among producers is successful, such that high quality is attained through collective production, I consider the role that different levels of expertise in production, and the interplay between novice and expert contributors affect whether the good produced satisfies heterogeneous consumer needs for collective goods. Three main findings emerged. First, I have documented the existence of a misalignment between the production of high quality goods and consumer needs. Second, as predicted, I have found that novice contributions play an important role in signaling to expert producers which goods are

most needed by consumers. In doing so, I have identified a social mechanism that links the micro-sociological dynamics of collective production to the resultant market-level macro-sociological order. Additionally, I have shown that expert contributors are motivated by responding to consumer needs, and that they interpret novice contributions as a cue that indicates which goods are most needed. Taken together, these findings offer several theoretical contributions: they draw attention to the potential misalignment between collective production and consumption of public goods, highlight a social mechanism through which individual level behavior results in macro-level outcomes, and shed light on the importance of consumer cues for the production of needed goods in collective production markets.

While the proposed theory is empirically tested in the context of the online encyclopedia Wikipedia, heterogeneous collective goods are a frequent occurrence in the contemporary life of markets, organizations and communities. These goods belong to one of three main categories: material goods such as bridges, or clean cities; information goods, such as open source software or knowledge databases; and norm-based goods, such as the content or enforcement of norms, rules and regulations that serve in preserving social order. There are two main criteria in identifying relevant contexts for the proposed mechanism: one, the public good produced is a heterogeneous public good that affects a large number of actors, and, two, while anyone may participate, one requires a high-resource investment in the production process to develop expertise and make high-quality contributions. Communication and knowledge infrastructures in organizations (Monge and Contractor, 2003; Tang, 2008), rules and regulations in an organization, industry, or country (Drahos, 2004), and basic research, open source software (Osterloh and Rota, 2007), and online collective production platforms such as consumer review

websites are all heterogeneous public goods meeting the above criteria, in the sense that different stakeholders are interested in different parts of the good, or in different attributes.

Collective Production Forms and Misalignment

In recent years, organizational scholars have focused on various forms of collective production as a context for bringing to the fore old and new organizational theories regarding categories (Leung, 2014), status and reputation (Kovács and Sharkey, 2013), social rewards (Piskorski and Gorbatai, 2010), and identity (Willer, Flynn and Zak, 2012). This research has considered motivation for participating in the collective production, but has not examined how motivation affects the relationship between collective goods produced and the needs of the consumers, or the dynamics among different types of participants in production process.

Building on these insights, I advance the organizational theory literature by formalizing how collective production enables coordination and collaboration among large numbers of actors working towards a shared purpose. Additionally, I highlight a social mechanism that accounts for increased alignment between the production and consumption of these goods. I show that novice contributors' participation has a direct negative effect on the quality of goods produced, but a positive indirect effect because it acts as a cue for expert contributors to improve the quality of those goods that consumers are most interested in. This conclusion serves to highlight the value of attending to the social interactions among actors with different levels of expertise engaged in the pursuit of a common goal. While ostensibly novice contributors are, on average, not effective at improving collectively produced goods, and are not skilled enough to coordinate their work with expert contributors, they play an important role in the success of the collective production process. A sole focus on the actions and interactions of expert producers—who invest a lot of

effort in coordinating with one another to improve good quality, and account for the vast majority of contributions to collective production—would have missed the critical role of novice contributors. Only by recognizing that collective production occurs through the interplay between the two groups could I account for the unexpected role of novice contributors, and for the observed improvement in collective good quality, which in turns has implications for the sociology of markets.

Micro-interactions and the Sociology of Markets

The demand side has been relatively ignored in the sociology of markets (White, 1981; Fligstein, 2001). As early as 1981, White proposed that the aggregate volume and prices are of less interest than understanding "how terms of trade [among producers and consumers] establish themselves across differentiated products so as to give meaning to aggregate terms like supply and demand." This study examines how "terms of trade" among producers and consumers unfold in a special case, that of collective goods. In the case of markets for such goods, prices and volume are not part of the discussion: any public good is free and non-rival, such that theoretically an unlimited number of individuals may consume it without affecting anyone else's use of the good. Building on White's (1981), I examine the social mechanism (Hedström and Swedberg, 1998) through which micro-level interactions between experts producing collective goods, and the consumers of these goods, visible to producers only vicariously through novice activity, aggregate to market-level outcomes in terms of the distribution and quality of goods produced.

Additionally, this study innovates by examining a market for heterogeneous public goods. Many public goods exhibit this property; heterogeneity in consumer needs for public goods is particularly evident in the case of information goods, a type of good increasingly

prevalent in modern organizations, markets, and society (Zammuto et al., 2007; Castells, 2011). By doing so, I demonstrate the importance of re-visiting existing market theories in the context of information goods, and re-examining the dynamics of production in this setting. Assuming general interest in the production of collective goods as a whole, and examining the extent to which expert producers have created high quality goods would have obscured a core problem of this market: the scarcity of information about consumer needs. This would have failed to uncover the micro-level phenomena that I have observed. Only by examining good-level production dynamics from the perspective of the information held by producers could I identify the social mechanism through which producers receive cues about demand.

Collective Production and User Innovation

This study contributes to the user innovation literature by examining the importance of the relationship between experts and novices— or lead users and all other users— as a source of information. The user innovation literature is predicated on the idea that lead users are expert producers who are privy to information about demand for a particular product or feature (von Hippel, 1994; 1998). Lead users apply this sticky, local knowledge to create goods that they, and others like themselves, are interested in. This research paradigm has examined the governance mechanisms and participation norms among experts—lead users— as well as the relationship between incumbent firms and user innovator communities within an industry (West and Gallagher, 2006; von Hippel, 2007; West and Lakhani, 2008). The current study contributes to this literature by attending to the antecedents of user innovations—namely, the choice that expert producers make to develop one idea instead of another.

Here I show that, even in situations where no material rewards are expected, experts attend to cues from potential users, and I identify one example of such cue. This particular cue comes from users who are not "experts" but are motivated to participate in the collective production process, such that their attempts register as a sign of interest with expert producers. The proposed theory highlights the paradoxical role of novices in collective production—namely, the fact that they have a negative direct effect on the quality of goods produced, but induce a positive indirect effect through the signals they send to experts about interest in a particular good. Thus collective production is more likely to respond to consumer needs if two conditions are attained: there is a set of individuals willing to invest time and effort to develop expertise in this production process, and there are consumers who occasionally contribute to goods that are of interest to them.

Conclusion

In our global age, technology has enabled collective production platforms where large groups, from thousands to millions of individuals participate in the production of information-type public goods such as book, restaurant or movie reviews; open source software; and encyclopedic, travel or medical knowledge. These goods are freely available to anyone in the world, provided that they have the skills and resources to consume them. While technology lends these goods unprecedented scale and reach, it also poses concerns as it stands to magnify risks and pitfalls inherent in collective production processes, such as coordination costs and lack of information about consumer needs.

This study highlights an important but understudied aspect of the collective production process: the paradox of novice contributions. In so doing, this paper speaks not only to the

emerging literature on collective production forms but also to a broader literature on micro-sociological mechanisms in markets. Theorists have called for an exploration of how markets emerge, and thus far most answers to this question have focused on the historical emergence of traditional markets or on post-communist transition economies (Padgett and Powell, 2012). In this study I emphasize a different type of market—for heterogeneous public goods—and a social mechanism by which needed goods are produced in this market. It is my hope that this study will stimulate more research into non-price mechanisms that connect producers and consumers in collective production markets, and beyond.

Table 1. Logistic Regression Estimates Predicting the Quality of Article k and Negative Binomial Estimates Predicting Article k Length (Hypothesis 1)

	Article quality$_k$			Article length$_k$		
Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Cumul. expert$_k$	1.673***	1.565***	1.321***	0.409***	0.385***	0.394***
	(0.093)	(0.099)	(0.143)	(0.018)	(0.020)	(0.026)
Cumul. novice$_k$	-0.321***	-0.325***	-0.223*	0.131***	0.124***	0.179***
	(0.067)	(0.071)	(0.096)	(0.018)	(0.019)	(0.017)
Avg.views.$_k$ (AV_k)	-0.004	-0.002	0.079	0.014	0.015*	0.051***
	(0.028)	(0.028)	(0.067)	(0.008)	(0.008)	(0.013)
Length$_k$			0.587***			
			(0.123)			
Projects$_k$ (P_k)	-0.097	-0.101	-0.135	0.035**	0.023*	0.077*
	(0.053)	(0.052)	(0.146)	(0.011)	(0.011)	(0.032)
Importance$_k$ (I_k)		0.393***	0.502**		0.132***	0.190***
		(0.080)	(0.169)		(0.025)	(0.042)
Experts$_k$			0.006			0.007***
			(0.004)			(0.001)
Editors50%$_k$			-0.023			-0.035***
			(0.012)			(0.003)
Monitors$_k$(M_k)		0.139	0.366*		0.074***	0.169***
		(0.096)	(0.167)		(0.019)	(0.037)
Undo's$_k$		-0.003**	-0.007*		0.000	-0.004***
		(0.001)	(0.003)		(0.001)	(0.001)
Protected$_k$		0.023	0.048		-0.006	-0.016*
		(0.043)	(0.062)		(0.009)	(0.007)
AV_k* M_k (/10)			-0.207			-0.130**
			(0.171)			(0.047)
AV_k* P_k (/10)			0.122			-0.073
			(0.242)			(0.062)
AV_k* I_k (/10)			-0.194			-0.125*
			(0.237)			(0.059)
Category$_k$	No	No	Yes	No	No	Yes
Deg. Freedom	5	11	38	5	11	36
-Log (pseudo) likelihood	1,205.70	1,173.29	765.70	80,519.01	80,416.10	52,308.43

Huber-White robust standard errors in parentheses. * p<0.05, ** p<0.01, *** p<.001 (two-tailed tests). Constant term, article age, and controls for missing data on average views, monitors, importance and article length were omitted from the table

Table 2. Negative Binomial Panel Estimates Predicting Expert Actors' Edits on Article k during Interval t+1 (Hypotheses 2 and 3)

	Expert contributions$_{k\ t+1}$			
	Fixed effects		Random effects	
Variables	Model 7	Model 8	Model 9	Model 10
Novice edits$_{k,t}$ (log)		0.061***	0.244***	0.185***
		(0.008)	(0.006)	(0.006)
Page views$_{k,t}$	0.146***	-0.002	0.002	0.004
($V_{k,t}$)	(0.007)	(0.003)	(0.003)	(0.003)
Cumul. novice edits$_{k,t}$	0.100***	0.192***	0.061***	0.044***
	(0.024)	(0.016)	(0.003)	(0.003)
Cumul. expert edits$_{k,t}$	-0.190***	-1.446***	0.088***	0.063***
	(0.031)	(0.024)	(0.004)	(0.004)
Protected$_{k,t}$		0.060***		0.004
		(0.011)		(0.003)
Ratio minor$_{k,t}$		-0.014		-0.057***
		(0.013)		(0.012)
No edits$_{k,t}$		0.083***		-0.050***
		(0.009)		(0.008)
Experts$_{k,t}$		0.002***		0.002***
		(0.000)		(0.000)
Importance$_k$ (I_k)			0.040***	0.029***
			(0.007)	(0.007)
Projects$_k$ (P_k)			0.003	0.004
			(0.003)	(0.003)
Monitors$_k$ (M_k)			0.032***	0.039***
			(0.006)	(0.005)
$V_{k,t}$ * I_k (/10)			0.033***	0.024*
			(0.010)	(0.010)
$V_{k,t}$ * P_k (/10)			0.049***	0.017
			(0.010)	(0.010)
$V_{k,t}$ * M_k (/10)			0.071***	0.017*
			(0.008)	(0.008)
Category$_k$	No	No	Yes	Yes
Interval$_t$	No	Yes	No	Yes
Degrees of freedom	4	17	37	49
- Log likelihood	136,028.1	135,159.2	195,578.9	194,320.6

Standard errors in parentheses. * p<0.05, ** p<0.01, *** p<.001. Constant term, controls for views, importance, and monitors missing, ratio no comment and first time edits omitted from table.

REFERENCES

Andreoni, J.

1995 "Cooperation in public-goods experiments: Kindness or confusion?" The American Economic Review: 891-904.

Anthony, D., S. W. Smith and T. Williamson.

2007 "The quality of open source production: Zealots and good samaritans in the case of wikipedia." Rationality and Society.

Ariely, D., A. Bracha and S. Meier.

2009 "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially." The American Economic Review, 99(1): 544-555.

Borgatti, S. P. and R. Cross.

2003 "A relational view of information seeking and learning in social networks." Management Science, 49(4): 432-445.

Brown, J. S. and P. Duguid.

1991 "Organizational learning and communities-of-practice: Toward a unified view of learning, and innovation." Organization Science, 2: 40-57.

Butler, B., E. Joyce and J. Pike.

2008 "Don't look now, but we've created a bureaucracy: The nature and roles of policies and rules in wikipedia." SIGCHI Conference Proceedings.

Castells, M.

2011 The rise of the network society: The information age: Economy, society, and culture, Oxford: John Wiley & Sons.

Chaudhuri, A.

2011 "Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature." Experimental Economics, 14(1): 47-83.

Ciffolilli, A.

2003 "Phantom authority, self-selective recruitment and retention of members in virtual communities: The case of wikipedia." first monday, 8(12).

Coleman, J. S.

1988 "Free riders and zealots: The role of social networks." Sociological Theory, 6(1): 52-57.

Coleman, J. S.

1990 Foundations of social theory. Cambridge, MA, Harvard University Press.

Dahlander, L. and S. O'Mahony.

2010 "Progressing to the center: Coordinating project work." Organization Science, 22(4): 961-979.

Drahos, P.

2004 "The regulation of public goods." Journal of International Economic Law, 7(2): 321-339.

Fernandez, R. M. and D. McAdam.

1988 "Social networks and social movements: Multiorganizational fields and recruitment to mississippi freedom summer." Sociological Forum, 3(3): 357-382.

Fischbacher, U. and S. Gächter.

2008 "Social preferences, beliefs, and the dynamics of free riding in public good experiments." CESifo Working Paper Series No. 2491

Fligstein, N.

2001 The architecture of markets: An economic sociology of twenty-first century capitalist societies, Princeton, NJ: University Press Princeton.

Gambetta, D.

2009 Signaling. The oxford handbook of analytical sociology. P. Hedström and P. Bearman. Cambridge, Oxford University Press**:** 168-194.

Gerhards, J. and D. Rucht.

1992 "Mesomobilization: Organizing and framing in two protest campaigns in west germany." American Journal of Sociology, 98(3): 555-596.

Goode, W. J.

1979 The celebration of heroes: Prestige as a social control system. Berkeley, CA, University of California Press.

Gould, R. V.

1991 "Multiple networks and mobilization in the paris commune, 1871." American Sociological Review, 56(6): 716-729.

Grant, A. M. and F. Gino.

2010 "A little thanks goes a long way: Explaining why gratitude expressions motivate prosocial behavior." Journal of personality and social psychology 98(6): 946-955.

Hardin, R.

1982 Collective action. Baltimore, MD, Johns Hopkins University Press.

Hausknecht, M.

1962 The joiners: A sociological description of voluntary association membership in the united states. New York, Bedminster Press.

Hausman, J. A., B. H. Hall and Z. Griliches.

1984 "Econometric models for count data with an application to the patents - r & d relationship." Econometrica, 52: 909-938.

Haveman, H. A., H. Rao and S. Paruchuri.

2007 "The winds of change: The progressive movement and the bureaucratization of thrift." American Sociological Review, 72(1): 117-142.

Heckathorn, D. D.

1993 "Collective action and group heterogeneity: Voluntary provision versus selective incentives." American Sociological Review, 58: 329-350.

Hedström, P. and R. Swedberg.

1998 Social mechanisms: An analytical approach to social theory, Cambridge, UK: Cambridge University Press.

Holcombe, R. G.

1989 "The median voter model in public choice theory." Public Choice, 61(2): 115-125.

Howison, J., J. F. Olson, A. Kittur and K. M. Carley.

2011 "Motivation through visibility in open contribution systems." CMU working paper.

Johansen, L. H.

2002 "Transferable training as a collective good." European sociological review, 18(3): 301-314.

Kittur, A. and R. E. Kraut.

2008 "Harnessing the wisdom of crowds in wikipedia: Quality through coordination." Proceedings of the 2008 ACM conference on Computer Supported Cooperative Work.

Kollock, P.

1998 "Social dilemmas: The anatomy of cooperation." Annual review of sociology, 24: 183-214.

Kollock, P.

1999 The economies of online cooperation. New York, Routledge.

Kovács, B. and A. Sharkey.

2013 "The paradox of publicity: How awards can negatively impact the evaluation of quality." Forthcoming in Administrative Science Quarterly.

Kriplean, T., I. Beschastnikh and D. W. McDonald.

2008 "Articulations of wikiwork: Uncovering valued work in wikipedia through barnstars." Proceedings of the 2008 ACM conference on Computer Supported Cooperative Work.

Lakhani, K. R. and R. Wolf.

2005 Why hackers do what they do: Understanding motivation and effort in free/open source software projects. Perspectives on free and open source software. J. Feller, B. Fitzgerald, S. Hissam and K. R. Lakhani. Cambridge, MA, MIT Press.

Lampel, J. and A. Bhalla.

2007 "The role of status seeking in online communities: Giving the gift of experience." Journal of Computer-Mediated Communication, 12(2): 434-455.

Leung, M. D.

2014 "Dilettante or renaissance person? How the order of job experiences affects hiring in an external labor market." American Sociological Review, 79(1): 136-158.

Marwell, G., P. Oliver and R. Prahl.

1988 "Social networks and collective action: A theory of the critical mass. Iii." The American Journal of Sociology, 94(3): 502-534.

McCarthy, J. D. and M. N. Zald.

1973 The trend of social movements in america: Professionalization and resource mobilization. Morristown, NJ, General Learning Press.

McElroy, G.

2006 "Committee representation in the european parliament." European Union Politics, 7(1): 5-29.

Mockus, A., R. T. Fielding and J. Herbsleb.

2000 A case study of open source software development: The apache server. Proceedings of the 2000 International Conference on Software Engineering, IEEE.

Mockus, A., R. T. Fielding and J. D. Herbsleb.

2005 Two case studies of open source software development: Apache and mozilla. Perspectives on free and open source software. J. Feller, B. Fitzgerald, S. Hissam and K. R. Lakhani. Cambridge, MA, MIT Press**:** 163-210.

Monge, P. R. and N. S. Contractor.

2003 Theories of communication networks. New York, Oxford University Press.

Musick, M. A. and J. Wilson.

2007 Volunteers: A social profile. Bloomington, IN, Indiana University Press.

Nov, O.

2007 "What motivates wikipedians?" Communications of the ACM, 50(11): 60-64.

O'Mahony, S. and F. Ferraro.

2007 "The emergence of governance in an open source community." Academy of Management Journal, 50(5): 1079-1106.

Oberschall, A. R.

1994 "Rational choice in collective protests." Rationality and Society, 6(1): 79-100.

Oliver, P.

1984 "If you don't do it, nobody else will: Active and token contributors to local collective action." American Sociological Review, 49(5): 601-610.

Oliver, P. and G. Marwell.

1992 Mobilizing technologies for collective action. Frontiers of social movement theory. A. Morris and C. Mueller. New Haven, CT, Yale University Press.

Olson, M.

1965 The logic of collective action, Cambridge, MA, Harvard University Press.

Osterloh, M. and S. Rota.

2007 "Open source software development—just another case of collective invention?" Research Policy, 36(2): 157-171.

Oz, A.

2009 "'Move along now, nothing to see here': The private discussion spheres of wikipedia." Available at SSRN 1726450.

Padgett, J. F. and W. W. Powell.

2012 The emergence of organizations and markets, Princeton, NJ: Princeton University Press.

Panciera, K., A. Halfaker and L. Terveen.

2009 Wikipedians are born, not made: A study of power editors on wikipedia. Proceedings of the ACM 2009 international conference on Supporting Group Work, ACM.

Piskorski, M. J. and A. Gorbatai.

2010 Testing coleman's social-norm enforcement mechanism: Evidence from wikipedia, Harvard Business School.

Ransbotham, S., G. C. Kane and N. H. Lurie.

2012 "Network characteristics and the value of collaborative user-generated content." Marketing Science, 31(3): 387-405.

Restivo, M. and A. van de Rijt.

2012 "Experimental study of informal rewards in peer production." PloS one, 7(3): e34358.

Schradie, J.

2011 "The digital production gap: The digital divide and web 2.0 collide." Poetics, 39(2): 145-168.

Schroer, J. and G. Hertel.

2007 "Voluntary engagement in an open web-based encyclopedia: Wikipedians, and why they do it." Proc. Congress of the European Association of Work and Organizational Psychology.

Setia, P., B. Rajagopalan, V. Sambamurthy and R. Calantone.

2012 "How peripheral developers contribute to open-source software development." Information Systems Research, 23(1): 144-163.

Shah, S. K.

2006 "Motivation, governance, and the viability of hybrid forms in open source software development." Management Science, 52(7): 1000-1014.

Shaw, A.

2012 "Centralized and decentralized gatekeeping in an open online collective." Politics & Society, 40(3): 349-388.

Snow, D. A., L. A. Zurcher and S. Ekland-Olson.

1980 "Social networks and social movements: A microstructural approach to differential recruitment." American Sociological Review, 45: 787-801.

Soetevent, A. R.

2005 "Anonymity in giving in a natural context—a field experiment in 30 churches." Journal of Public Economics, 89(11): 2301-2323.

Sun, M. and F. Zhu.

2012 "Ad revenue and content commercialization: Evidence from blogs." Available at SSRN 1735696.

Tang, L.

2008 "An integral model of collective action in organizations and beyond." Journal of business ethics, 80(2): 249-261.

Traxler, F.

1993 "Business associations and labor unions in comparison: Theoretical perspectives and empirical findings on social class, collective action and associational organizability." British Journal of Sociology, 44(4): 673-691.

von Hippel, E.

1994 "'Sticky information' and the locus of problem solving: Implications for innovation." Management Science, 40(4): 429-439.

von Hippel, E.

1998 "Economics of product development by users: The impact of “sticky” local information." Management Science, 44(5): 629-644.

von Hippel, E.

2007 "Horizontal innovation networks—by and for users." Industrial and corporate change, 16(2): 293-315.

von Krogh, G., S. Spaeth and K. R. Lakhani.

2003 "Community, joining, and specialization in open source software innovation: A case study." Research Policy, 32(7): 1217-1241.

West, J. and S. Gallagher.

2006 "Challenges of open innovation: The paradox of firm investment in open-source software." R&D Management, 36(3): 319-331.

West, J. and K. R. Lakhani.

2008 "Getting clear about communities in open innovation." Industry and Innovation, 15(2): 223-231.

White, H. C.

1981 "Where do markets come from?" American Journal of Sociology: 517-547.

Whitmeyer, J. M.

2007 "Prestige from the provision of collective goods." Social Forces, 85(4): 1765-1786.

Willer, R.

2009 "Groups reward individual sacrifice: The status solution to the collective action problem." American Sociological Review, 74(1): 23.

Willer, R., F. J. Flynn and S. Zak.

2012 "Structure, identity, and solidarity: A comparative field study of generalized and direct exchange." Administrative Science Quarterly, 57(1): 119-155.

Yang, H.-L. and C.-Y. Lai.

2010 "Motivations of wikipedia content contributors." Computers in Human Behavior, 26(6): 1377-1383.

Zammuto, R. F., T. L. Griffith, A. Majchrzak, D. J. Dougherty and S. Faraj.

2007 "Information technology and the changing fabric of organization." Organization Science, 18(5): 749-762.