ALIGNING COLLECTIVE PRODUCTION WITH DEMAND:

EVIDENCE FROM WIKIPEDIA*

Andreea Gorbatai

Harvard University

September 12, 2011

Keywords: collective goods, non-price alignment mechanism, collective production, expertise, production, collective action

Word Count: 15,919

* **Acknowledgements:** I am grateful for the most helpful comments and suggestions to Mikolaj Jan Piskorski, Richard Hackman, Peter Marsden, Michel Anteby, Sameer Srivastava, Andras Tilcsik; to the participants in the Work, Organizations and Markets workshop at Harvard Business School; and to 2011 Sorensen/ Inter-Ivy Conference participants. **Corresponding Author: Andreea Gorbatai,** Cotting 324A, Harvard Business School, Soldiers Field, Boston, MA 02163. Email: agorbatai@hbs.edu.

Abstract

Economic markets align supply and demand through prices. However, many social phenomena lack pricing to inform producers about consumer demand. This can lead to the over- or under-production of certain goods and services. In this paper, I propose a social mechanism that aligns collective production with demand. I argue that this will occur as long as three conditions are met: first, that consumers attempt to become occasional producers; second, that these attempts are observed by producers who interpret them as signs of unmet demand; and lastly, that producers are willing and able to provide these goods. I test this theory using a large dataset of English-language Wikipedia articles, created by merging information from 185 million records of article contributions with data on article views, quality, and knowledge type. The results are consistent with my theory, suggesting that this social mechanism can take the role of prices in certain markets. This study has implications for the sociology of markets, as it highlights a largely ignored collective action-demand alignment mechanism, and for collective production in organizational and non-organizational settings such as collective creativity, scientific collaboration, and community production of collective goods, as it identifies an unexpected benefit from consumer contributions.

ALIGNING COLLECTIVE PRODUCTION WITH DEMAND: EVIDENCE FROM WIKIPEDIA

Social scientists have long studied mechanisms that align production with demand (Swedberg 2005). Economic theory suggests that resources are allocated most efficiently through the price mechanism (Coase 1937; Eckstein and Fromm 1968). Failures of this mechanism due to information asymmetries, externalities, or market power may lead to the under-provision of important goods (Akerlof 1970). In such circumstances, firms (Coase 1937; Williamson 1975), the government, or embeddedness in social networks (Granovetter 1985; Uzzi 1996) can help improve resource allocation (Fligstein 2001; Weisbrod 1986).

Although social scientists have examined the alignment of production and consumption in economic markets (White 1981, 2002), less is known about the alignment of the demand for and the provision of collective goods, defined as goods made available for consumption by all members of a group whenever they are made available for consumption by any one of them (Olson 1965). For example, the literature on collective action has studied the conditions under which social movements come into being, but has not examined whether the social movements most demanded by the people come into being (McAdam, Tarrow, and Tilly 2001; Tarrow 1994). Similarly, literature on volunteer work rarely examines whether the most-needed causes receive the most attention (Wilson 2000). If anything, research leads us to believe that participation in collective good production is driven by internal motivation or selective incentives such as peer rewards and rarely by demand for various volunteer actions (Knoke 1988; Kyriacou 2010; McAdam 1985; Oliver 1984; Shah 2000; Tarrow 1994; Whitmeyer 2007;

Baldassarri 2009; Willer 2009). In addition, even if contributors wished to provide collective goods to meet demand, they may lack clear and timely information about existing demand. Both of these mechanisms suggest that individuals may fail to produce the collective goods that satisfy existing demand. This could translate in misallocation of resources and in invisible and unanticipated inequalities in individual satisfaction with collective good provision, despite the fact that certain collective goods were successfully produced.

In this study I identify a previously overlooked mechanism for aligning collective action with demand for collective goods where no price mechanism exists. Specifically, I theorize that collective production can be brought in line with consumer demand for collective goods when three necessary conditions are met: i) consumers become involved as occasional producers of certain goods, ii) consumers' involvement is observed by producers who interpret their participation as unsatisfied demand, and iii) producers are willing and able to improve these collective goods.

To test this mechanism linking consumer demand with improved collective goods provision, I rely on one well-known example of collective production: the free online encyclopedia Wikipedia. I employ a unique panel dataset created by merging five different data sources to examine the conditions under which the volume and quality of over three million Wikipedia articles produced through 185 million contributions is related to revealed demand for these articles, as measured by article page views across time. This dataset is uniquely suited to address the research question because it contains longitudinal data on demand for collective goods that is available to the researcher but not to the producers of these goods—article contributors cannot observe article page views—such that we can examine the social mechanisms which lead to appropriate goods provision when producers lack direct information about demand. This dataset also contains information on collective good (article) quality, which is often difficult to procure in offline settings and difficult to assess objectively when comparing other collective goods (e.g., different amenities of a public area or different political demands). In addition to information on article production, demand, and quality, I include data about the knowledge category of each article and about the number of producers monitoring each article for changes in order to control for inherent differences in the type of good and in the likelihood that producers would observe changes to a good. Before discussing my dataset and empirical findings, I review sociological theories of economic production and markets, and propose to test the existence of a mechanism that connects consumers and producers of collective goods.

THEORY: ALIGNMENT MECHANISMS

The Alignment of Economic Production to Demand

Economists and sociologists have devoted considerable attention to the question of how good production aligns with consumption across various market settings (for a review, see Aspers 2009). Most of this research assumes that "the most fundamental feature of the economic system [is] *production for* [*the*] *market*" (Knight 1971;White 2002) - the end consumer or market intermediaries who can affect the ends desired by the organization. Central to the functioning of this system are the rules of exchange and resource scarcity, which represents the basis for valuation. Material goods, human resources, and financial capital are all characterized by scarcity because employing them in one setting entails opportunity costs. When a certain resource is scarce in a market, it will be priced at a premium relative to other resources and will play a more important role in market processes and structure. Specifically, goods commanding excess prices will attract new producers to the market, who will then bargain down the price until no producer finds it profitable to enter. At this point, the structure of production reflects the structure of underlying demand and no improvements to welfare can occur (Arrow 1977).

This idealized image of a well-functioning market has been subject to scrutiny by social scientists (Baker 1984; Beckert 1996; Granovetter 2005). Theorists have identified at least three conditions that prevent market prices from adequately representing demand, leading to market failures: the presence of externalities, market power, and information asymmetries (Weisbrod 1986).[1] For example, studies of economic markets have shown that status acts as a signal affecting firms' costs and revenues independent of product quality (Podolny 1993), and that firms with personal contacts to a banker benefit from lower interest rates (Uzzi 1999). The identification of such failures led to efforts to address them, by supplanting market mechanisms with hierarchies, hybrid organizations, or government regulations, all of which were intended to alter the production function (Weisbrod 1986). Sociologists have contributed to this literature by examining the role of hybrid organizations and networks (Gulati 1998; Gulati and Gargiulo 1999; Uzzi 1999) and also by studying producer organizations and regulating activity (Fligstein 2001; Zuckerman 1999).

The Alignment of Collective Production to Demand

Sociologists have paid relatively little attention so far to the alignment of production and consumption in collective action settings compared against alignment in economic markets, mainly because data on demand for collective action are difficult to procure. Consider, for example, the arena of collective action, which is defined as the pursuit of common goals by a set of persons who overcome incentives to free ride (Marwell and Oliver 1993; Olson 1965; Schelling 1973). A substantial body of sociological work either examines these phenomena from a theoretical collective action perspective (Marwell and Oliver 1993; Marwell, Oliver, and Prahl 1988; Oliver, Marwell, and Teixeira 1985) or focuses on the particularities of resource mobilization, opportunity structures, and political processes in the context of social movements (McAdam, Tarrow, and Tilly 2001; Tarrow 1994). Research explains how the production of a social movement takes place in relationship to other movements, to those in power, and to the broader public potentially interested in the goals of the movement (Giugni 1998), but does not address how the needs of beneficiaries are communicated to social movement participants, or whether those needs are represented among the social movement's goals.

Similar concerns abound with collective production endeavors, defined as collective action oriented towards the production of collective goods. When collective action succeeds at producing goods such as software (the open source movement), political or regulatory changes (social movements), or public spaces (neighborhood initiatives), the efforts are widely praised. However, little work has been done to assess to what extent, and through what means, these collectively produced goods meet actual demand.[2] Making this assessment is challenging because in absence of aggregate information such as pricing it is difficult for producers[3] to know what consumers need and whether these demands are satisfied.

A priori, there are a number of theoretical reasons to believe that there is significant misalignment between what is needed and what is provided. Since it is difficult for producers to respond directly to consumer demand, they may produce goods that are not needed while failing to produce others that are. While emergent economic markets reach a stabilization stage where the attributes of goods exchanged and the rules of the exchange are clear to all parties (Aspers 2009), collective production rarely progresses to a stage where participants converge on these factors, primarily because it does not take place for exchange purposes between producers and consumers. Collective producers may participate for intrinsic reasons, such as enjoyment in the process or creating something they want or need, or for altruistic reasons they value independent of personal benefits. For example, a collective producer may be motivated to help create a neighborhood park because she believes it is important to have a green recreational area even if local inhabitants do not express interest in such a space; a musician may play in public spaces without expecting financial reward simply because she wants to share her love of music; and a cook may share a healthy recipe for free because she believes it will benefit public health. As peer producer relations emerge or as members of one's personal network join a cause, the social structure of collaboration among producers motivates additional contributions through identity and status rewards and other social incentives (Coleman 1988; Zhang and Zhu 2011), which promote collaborative production and make participants feel they belong and are socially valued (Grant and Gino 2010).[4] This logic suggests that collective production participants may be unaware of consumer needs beyond their peer producers' and may fail to produce certain goods unless certain conditions are met.[5] This is in contrast to economic markets where the price mechanism and profit motive ensure that production aligns to consumer demand.

Identifying a Collective Production Mechanism

To explore the mechanism through which collective production participants may become aware of consumer demand for specific goods, I start with a simple model. I assume there are two types of actors, actor-producers in the collective production process, that I label 'expert producers,' and actor-consumers who stand to benefit from collective production outcomes but are not involved in production, that I label 'consumers.' While in practice expert producers are consumers of collective production outcomes as well, I assume that they represent a small percentage of a much larger population of consumers and are not necessarily a representative sample of consumers. This assumption is consistent with the literature on collective action that finds that not all potential beneficiaries of a cause contribute to it (Marwell and Oliver 1993). It parallels the distinction made in the social movements literature between activists who "care enough about some issue that they are prepared to incur significant costs and act to achieve their goals" (Oliver and Marwell 1992) and non-producers who may make small efforts or material contributions if asked directly by an activist/producer or "implicitly" provide support through an event that presents an occasion for decision-making at an individual level (Collins 1981). I also assume that the collective production system is open to participation, such that anyone can join; however, new participants will lack knowledge about how to effectively contribute to the collaborative production process. Lastly, I assume that the collaborative production process enables expert producers to observe novice participants' attempts to participate but not the act of consuming the collective good itself. In practice, these observations may occur through various mechanisms, depending on the nature of the collaborative environment and of the contribution.

The proposed mechanism works as follows. First, consumers who are seeking to consume goods that have been insufficiently provided express demand through an attempt to contribute (novice producers). Experts observe the novice production, interpret it as a sign of unsatisfied demand, and respond by increasing contributions towards the creation of the good in question. The resulting expert production leads to a higher quality and volume of needed goods (Figure 1). I explain this mechanism in more detail in the next section.

Insert Figure 1 about here

Consumer demand and novice participation. To examine the link between demand for collective goods and production, I start by considering consumer behavior. When demand is not satisfied by collectively produced goods, consumers may seek an alternative or, if they have the means, they may attempt to improve the goods themselves. For example, a consumer of a travel information site may notice a lack of visual information about a location and contribute a photograph of her trip there, but she may not know how to label it such that it easily benefits others. A consumer of pre-1950s mathematical tables may notice an inconsistency in the natural logarithms table and point it out to the human computers despite not knowing exactly how to correct it (Robson et al. 2003). An individual who is marginally involved in neighborhood affairs may report a negative situation to a neighborhood committee or local newspaper with the expectation that the appropriate authorities will look into it (Smith 1976). Fans of musical styles or other entertainment and cultural goods may show overt support for the goods produced by their favorite artists, consequently fueling improved provision of the most demanded goods (Whitmeyer 2007). Although collective action beneficiaries may be interested in a certain outcome, many still fail to deeply engage in production for a variety of reasons cited in the literature on free riding (Olson 1965) and social movement participation (Fernandez and McAdam 1988; Klandermans 2004; Passy and Giugni 2001). This observation suggests the following prediction regarding novice participation in collective production:

*Hypothesis 1***:** An increase in demand for a collectively produced good increases the likelihood that consumers will attempt to produce that good.

Novice participation as signaling to experts.[6] Research on how contributions are distributed among participants in online collective production settings reveals a pattern whereby a small number of contributors are prolific producers, while large numbers of consumers never produce or are occasional, novice producers (Wilkinson 2008; Wu, Wilkinson, and Huberman 2009). The fact that many participants are superficially involved resonates with broader findings regarding reasons for involvement or non-involvement in collective action and social movements. First, "the perceived effectiveness of one's own potential contribution" (Passy and Giugni 2001) or beliefs about the likelihood that others would participate (Oliver 1984) generate differential participation in collective action. Second, communication with peers is important for successful participation in collective action because it "enables a person to find out about others' choices, to make explicit commitments, to appeal to what is the moral thing to do, and most importantly, to create or reinforce a sense of group identity" and shared purpose (Kollock 1998). Hence, when individuals are interested in a collective good but expect that others will create it for them, they may stop participating before learning how to make effective contributions themselves or they may fail to communicate with peers about improvements in participation and outcomes. Thus, their levels of contribution and experience in collective action remain low despite their interest in the goods.

However, when individuals dedicate substantial time and effort to furthering the goal of the collective production project, they become expert producers. As such, they are cognizant of collaboration norms and contribution processes, as are for example social activists in social movements. These experts employ various collaboration channels and tools to monitor and improve the collective goods produced. But in the case of many collective goods, it is virtually impossible to aggregate consumer demand:[7] one cannot know whether people would react to political goods that are not on a movement's agenda, or whether people are dissatisfied with the state of a nature preserve. Hence I propose:

Hypothesis 2a: An increase in demand will have no effect on the number of contributions by expert producers.

Expert producers may, however, observe if individuals or social groups attempt to add new items to a political agenda, or attempt to improve the state of the natural preserve. I predict therefore:

Hypothesis 2b: An increase in production by occasional producers will lead to an increased number of contributions by expert producers.

Collective production outcomes. Increasing the volume of a good is easier than increasing its quality. For example, it is relatively easy to make contributions such as removing an unnecessary line from software code documentation, sending a message or signature of support, offering a financial donation to a cause, or participating in a peaceful street protest. Since contributing to collective goods by adding or removing features entails work that can be easily performed by novices and experts alike (as compared against integrating or synthesizing work) I would expect that both expert and novice contributions are positively related to the volume of collectively produced goods in the presence of very low coordination costs associated with increasing volume:

Hypothesis 3a: An increase in the number of contributions by expert producers will lead to an increase in the volume of goods produced.

Hypothesis 3b: An increase in the number of contributions by occasional producers will lead to an increase in the volume of goods produced.

Comparatively, more complex work is required to increase the quality of a collectively produced good. Production of higher-quality outcomes requires not only that producers add new information, but also that they meet production standards, respect collaboration norms, and seek and accept feedback and even contestation from peers (Okhuysen and Eisenhardt 2002). These actions require skills such as negotiating, synthesizing, or integrating new information with existing features by smoothing out contradictions, reducing redundancies, and/or communicating and coordinating with other producers (Okhuysen and Eisenhardt 2002). In contrast to novices, collective production experts are by definition much more apt to achieve these complex tasks in the context of a particular collective production project. I expect therefore that expert contributions positively impact the quality of produced goods, while novice contributions pose risks to quality:

Hypothesis 4a: An increase in the number of contributions by expert producers will lead to an increase in the quality of the goods produced.

Hypothesis 4b: An increase in the number of contributions by occasional producers will lead to reduced quality of the goods produced.

Given that consumption on its own does not translate into feedback or contribute to collectively produced goods, I would expect that, holding everything else constant, consuming collectively produced goods has no direct effect on outcome quality:

Hypothesis 4c: Increased demand will have no effect on the quality of the goods produced.

TESTING THE THEORY WITH ENGLISH WIKIPEDIA DATA

I test these hypotheses in the context of Wikipedia, a free online volunteer-contributed encyclopedia and a salient example of the collective production process. Because individual contributions are not censored or screened before being included in the encyclopedia,[8] Wikipedia has attracted over six million registered contributors who produced over 3.5 million articles in English and over 16 million articles total in over 260 languages by September 2010. As of this date, Wikipedia also registered approximately 477 million views per day, half of which were to English-language pages, making it the seventh most visited website in the world.[9] Many consumers of Wikipedia's content are not aware that they can contribute, and many of those who contribute do not create participant accounts, choosing instead to edit anonymously. Of those who create accounts, most do not make more than two or three contributions (ever) to articles; less than one in five registered editors contributes more than ten times.[10]

The success of Wikipedia relies on a technology called wiki software, which enables people to interact with formerly static website pages. Individuals can modify any existing page for everyone else to see, while previous versions of the page remain accessible on a history page (for a description of the relation between main pages and history pages, Table 1).[11] Participants disagreeing with changes made on a page may immediately alter or erase these changes. Since the contributions of various participants are meshed together in the article text, interdependence in the collaboration process is inherently reciprocal (Thompson 1967), and authorship is obscured from consumers who read the final product. This reciprocal interdependence and lack of individual ownership and control are common in collective action phenomena.

Insert Table 1 about here

Collective Production Rules and Expertise

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 norms (Table 2).[12] Even 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).

Insert Table 2 about here

Although Wikipedia remains in principle an encyclopedia that anyone can edit, 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). 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 new, [novice producers]," keeping novices 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 2008). Researchers have argued that such spaces enable Wikipedia expert members 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 2008).

"Experts" are therefore not necessarily knowledge experts but individuals who understand the contribution process and are privy to and often involved in Wikipedia's "private sphere." 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 collective production but rather a collection of pages and a source of information. Other forms of collective actions 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 [experienced participants]. [Moreover], it is not certain that they will contribute, even if they are asked" (Oliver and Marwell 1992).

Mapping the Proposed Mechanism

To explore the mechanism through which collective production participants may become aware of consumer demand for specific goods, I first engaged in participant observation of contribution to Wikipedia by engaging in Wikipedia edits and observing other contributors' editing work and discussion threads regarding both general policies and coordination in article writing. My observations were conducted between December 2006 and May 2011, during which time I also interviewed a random 35 expert contributors to Wikipedia.[13] Analyzing my expert producer interview data, I locate qualitative evidence for the proposed social mechanism connecting expert producers to latent demand as a result of novice production.

Consumer demand and novice participation. According to my interviewees, the majority of Wikipedia contributors started by chance, when they encountered missing or erroneous information in an article they were reading which they could easily correct. Hence, contributors come to collective production as novices by way of articles they consume;[14] this supports the first hypothesis that the more demand for an article, the more likely it is the article will receive contributions from novice producers. According to my interviewees, their 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.[15] 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 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). Given that repeated contribution leads to learning and socialization, I assume there is a positive relationship between the number of contributions to Wikipedia articles and expertise, with a large number of infrequent contributors never progressing beyond the consumer/novice-producer stage.

Novice participation as a signal to experts. Wikipedia expert producers have two means of observing changes to articles they are interested in. First, they can periodically revisit articles and examine the history page, which contains a reverse-chronological index of all modifications to that specific article. “I look at [articles I edited] a little bit ... I go back and I check maybe once every few weeks or so just to see what’s happened. I look at the history and see what, you know—what’s going on with that page,” explained one interviewee.

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 “watch list.” Many interviewees reported actively monitoring anywhere from hundreds to thousands of articles. In particular, they frequently visit articles that were modified by novices to check on their work, amend these contributions, respond to suggestions, or remove malicious or damaging work. One interviewee explained his strategy in dealing with the large volume of article contributions:

> When I log on to Wikipedia, ... I log onto my watch list as the very first thing [and] there [are] hundreds of things that have changed [on monitored articles], and what do I concentrate on? … *I look for articles that have been changed by occasional editors because I don't trust them yet because … some add good content, but many, many do not.* And there [are] editors whose names I recognize, and … I might ignore [their changes] completely even if I might not agree [because] I think, "Okay, this editor isn't a bad person…*I don't know what they did, and … I won't look at it*."

The interviewees clearly stated that novice contributions to articles act as a signal that readers were interested enough to have read the article attentively and find ways to improve it:

> I'm a fast writer, so I'll often miss punctuation or spelling errors, which other people will fix...It's neat seeing that [p]eople are reading [what I write], 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.*

> Some kinds of [anonymous novice] edits are good in themselves but do not conform to the encyclopedic style or to the coherence of the article. [Regardless, *things] like fixing my typo shows someone is reading and paying attention—that*

is motivating. Sometimes [anonymous novices] will post … "This aspect of the article needs more coverage" and sometimes they are completely right and that makes me … add the content they suggest.

These interviews suggest that although novice contributions to collective action may consist in small, inconsequential changes, their participation signals to expert producers that certain goods are of interest to consumers. In addition to signaling interest, novice contributions may create the opportunity for innovative and unexpected associations between the existing information and new information provided by the novice (Brown and Duguid 1991). An example of information that benefits from novice contributions is provided by a Wikipedia expert editor who explained his work improving foreign speakers' contributions to the English Wikipedia: "My focus changed gradually … to the history of Hong Kong … because … more and more scientists came in to write, but not as many Hongkongers who write well in English [so I help with editing their contributions]."

Collective production: Process and outcomes. My expert producer interviewees explained that as they continued to participate they started to gain expertise in making productive contributions to articles and gradually became aware of the presence of other contributors, learning to locate and communicate with them and to receive feedback on their work. Expert contributors agreed that even when they do not explicitly collaborate with each other, they are aware of other experts in their knowledge areas and consult each other on specific issues. Several interviewees explained:

> If I'm editing Chinese cuisine articles, *there are two editors that are very knowledgeable* about that. And one, I think [he's] Chinese-Canadian [is] really knowledgeable about Chinese esoteric ingredients, particularly … sauces … so if I'm interested in something and I don't know about it, he can read Chinese, so then I'll [leave a message], "You know, I'm working on this article—[can you] look up this term?"

> I am one of ten editors who work extensively on [weather] articles. Occasionally I ask for help, but more often I work alone. However, *we regularly provide each other help* when it comes to copyediting/final touches.

> I took the lead in improving [certain articles] because I had the necessary expertise and the articles were extremely deficient [but] then *if someone else with expertise shows up it does entail a dialogue*… I did most of the writing [on these articles] and [others] worked more as reviewers improving writing and my use of linguistic terminology etc.

The experts interviewed suggested that, aside from periodic consultations, they do not significantly interfere with articles that someone else has taken the lead on, out of respect for their expert peers. One interviewee explained: "One thing I try to do [when I contribute to an existing article is that] I never change the structure that … is there, I will try to keep it … not to be too arrogant. A contributor is obligated to those before to maintain the work and add to it." Another explained that his work with other experts mainly consists in discussing the wider knowledge topic and reviewing each other's writing, but otherwise not interfering because his peers "know how to write and get it good on the first attempt." In the next section, I briefly summarize why Wikipedia represents an ideal setting for testing the proposed mechanism.

The Fit between Theory and Empirical Setting

Wikipedia offers a number of advantages as a research setting for testing the theory that demand for goods influences production through novice involvement. First, by the nature of its online platform, Wikipedia collects an unprecedented amount of longitudinal data on both the production and views (consumption) of encyclopedic articles, while collective producers are unable to directly observe article views.[16] Second, we can assume that a Wikipedia article's page views reflect consumer demand, given that as of 2007 more than 96% of Wikipedia article links were ranked on the first page of Google search results. Third, encyclopedic articles are a type of good that requires minimal skills (i.e., literacy) to consume. Given that articles are easy to find and consume I can reasonably assume that article views approximate reading, which in turn approximates expressed preferences for a free good.

Lastly, Wikipedia's technological platform and collective production norms evolved in such a manner that it is relatively easy for consumers to make basic contributions but difficult for them to take advantage of the full array of options for contributing and collaborating. While this feature is not unique to Wikipedia, the digital traces of novice contributions are easier for researchers to identify than in other settings. Taken together, all these characteristics recommend Wikipedia as the ideal research setting for the proposed collective production theory. Relying on my access to data on article modification history, views, quality, length, and monitoring, I set out to test the proposed mechanism.

DATA AND METHODS

The comprehensive panel dataset employed in this study was created through the merger of several unique data streams provided by volunteer Wikipedia participants 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 (Table 3).[17] 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,[18] where an interval represents a half-month period for the production (article edits) and consumption (article views) data between October 2008 and January 2009.

Insert Table 3 about here

As I have argued above, expert contributors on Wikipedia are those who have experience collaboratively participating in the article production process. 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, where novices have a registered account or contribute anonymously.[19]

Models and Dependent Variables

I employ an article-interval analysis for the first two hypotheses, followed by a cross-sectional analysis of article quality and length, given that my dataset contains longitudinal data on article production and consumption, and cross-sectional data for article quality and length. Three of my four dependent variables—novice contributions, expert contributions, and article length—are count variables 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 for the first three of my four models (Hausman, Hall, and Griliches 1984). The variances of the first three dependent variables—novice contributions, expert contributions, and page length—are much greater than their means, which is indicative of overdispersion, so I assume a negative binomial distribution. The general log-likelihood function for binomial models is:

$$L(\beta \mid y, X) = \prod_{k=1}^{N} \Pr(y_k \mid x_k) = \prod_{k=1}^{N} \frac{\Gamma(y+\alpha^{-1})}{y!\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_k} \right)^{\alpha^{-1}} \left(\frac{\mu_k}{\alpha^{-1}+\mu_k} \right)^{y_k} \qquad (1)$$

where μ, the expected value of the distribution, and α, the over-dispersion parameter, are the negative binomial distribution parameters; X is a vector of independent variables; and y is the dependent variable. I use both the fixed-effects and random-effects estimator proposed by Hausman et al. (1984), where the latter estimator assumes that over-dispersion due to unobserved heterogeneity is randomly distributed across articles, in order to use my time-invariant data.[20] I verify the robustness of my findings by examining consistency across these estimates. I also check these results against a quasi-maximum likelihood Poisson (PQML) estimator, which makes no assumptions about the distribution of the data and yields consistent coefficient estimates as long as the mean of the data is correctly specified, and consistent robust standard errors even if the mean is incorrectly specified (Silva and Tenreyro 2006; Wooldridge 2006).

Hypotheses 1, 2A, and 2B. In order to test Hypothesis 1, I model the number of contributions by novice editors to article k during time interval t as a function of article consumption during time interval t, as well as time-variant and invariant article characteristics. In order to test Hypotheses 2a and 2b, I model the number of edits by expert editors to article k during time interval (t+1) as a function of article consumption and novice editing during time interval t, the state of article completion as of time t, and time-variant and invariant characteristics of the article.

I first estimate within-group, fixed-effects negative binomial regression models, such that any possible source of variation across articles is controlled for. This strategy has the advantage of considering only within-article variance in the estimation of regression coefficients, so that the measured effect of consumption is independent of any unobserved attributes of the article. Fixed-effects estimators are often preferred because of the likelihood that the stronger assumptions behind the GLS estimator are not satisfied, implying poor finite sample properties (Angrist and Pischke 2009). The fixed-effects negative binomial model for panel data is a generalized form of the Poisson model where an individual unobserved effect χ_k in equation(2) and, respectively, δ_k in equation (3) is introduced in the conditional mean (Greene 2000):

$$E(novice\_edits_{kt} \mid x_{kt}) = e^{\varphi_0 + \text{consumption}_{kt} * \varphi_1 + X_{kt}\beta + \chi_k + \varepsilon_k} \quad \text{(Hypothesis 1)} \qquad (2)$$

$$E(expert\_edits_{kt+1} \mid x_{kt+1}) = e^{\eta_0 + \text{consumption}_{kt} * \eta_1 + \text{novice\_edits}_{kt} * \eta_2 + X_{kt}\phi + \delta_k + \varepsilon_k} \quad \text{(Hypotheses 2a,b)} \qquad (3)$$

where ψ_0, ψ_1, β, and respectively η_0, η_1, η_2, and φ are vectors of coefficients to be estimated.

However, because fixed-effects models ignore between-groups differences, this estimator is less efficient and does not use time-invariant data (Greene 2000). I check the robustness of my results using random-effects estimators, under the assumption that unobserved heterogeneity is uncorrelated with the regressors. Because the Hausman (1978) test for choosing between the two types of effects models was inconclusive—the difference in coefficients was not systematic—I present both types. Standard errors are based on the observed information matrix.

The choice of testing Hypothesis 1 by estimating novice contributions during the period of measured article consumption is informed by both collective action theory about novice participation and qualitative information about Wikipedia contributors. Non-participant actors are more likely to become involved in collective action when they are presented with the opportunity to contribute (e.g. requests to donate money or time or to endorse a cause); Wikipedia interviewees reported that their first contributions were similarly opportunistic (i.e. they were often the result of consuming an article and observing an opportunity to make a low-cost contribution). The choice of testing Hypotheses 2a and 2b by estimating expert contributions in the period following novice contribution is based on the rationale that contributions which closely follow 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 was made, the half-month interval seems sufficient to capture the long-term effects of novice participation on expert contributions.

Hypotheses 3A and 3B. In order to test Hypotheses 3a and 3b I study the relationship between article length at the end of the last interval and prior expert and novice contributions, both log-transformed. Article length does not equate to total effort because a short article may have taken a long time to collaboratively produce and copy-edit, while a longer article may have been easier to write. However, length measured as the number of characters represents a reasonable metric of the volume of information a consumer receives on a particular subject, although it may be difficult to determine the comprehensiveness of the information based on length alone. The cross-sectional negative binomial model employed in testing Hypotheses 3a and 3b is:

$$E(length_k \mid x_k) = e^{\lambda_0 + \text{avg.consumption}_k * \gamma_0 + \text{cumul.novice\_edits}_k * \gamma_1 + \text{cumul.expert\_edits}_k * \gamma_2 + X_k \vartheta + \varepsilon_k} \qquad (4)$$

where ε_k represents the error term, and λ_0, γ_0, γ_1, γ_2 and β coefficients to be estimated.

Hypotheses 4A, 4B, and 4C. Finally, I test Hypotheses 4a, 4b, and 4c using a logistic regression model to evaluate the extent to which expert contributions, novice contributions, and average article consumption, all three log-transformed, predict the quality of the good 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,[21] in which a set of participants interested in a broader subject related to the article's topic evaluate existing articles on that topic and coordinate 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 the article is likely to satisfy customer need, where articles with quality of one meet a minimum requirement (B-class or 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.[22] It is important to note that article maturity or age, measured as time since the article was created, is only weakly correlated with article quality, such that older articles are not always high quality, and newer articles are not necessarily low quality. The last equation takes the form:

$$\text{Quality} = f\left(z_k\right) = \left(1 + e^{\lambda_0 + \text{avg.consumption}_k * \gamma_0 + \text{cumul.novice\_edits}_k * \gamma_1 + \text{cumul.expert\_edits}_k * \gamma_2 + X_k * \gamma + \varepsilon_k}\right)^{-1} \tag{5}$$

where coefficients λ0, γ0, γ1, γ2 and the vector of coefficients γ are to be estimated.

In order to evaluate the robustness of these results and make use of the entire range of quality evaluations, I also test these hypotheses using the raw quality measures that range between 1 and 7 as dependent variables in an ordered logit regression. In the next section I explain my independent variable definitions, and then examine the results of the estimations, followed by a discussion on the limitations 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 Top10 (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.[23] 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-transformed. The other two important variables*, the number of novice edits* and *of expert edits*, were log-transformed for the same reasons when used as regressors.

Time Variant Control Variables

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.[24] 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's was included in some models; undo edits, which they are simply erasures, contribute less to article quality and length than other edits. Two other article characteristics controlled for are *ratio minor* and *ratio no comment*. *Ratio minor* is calculated as the number of inconsequential contributions (such as a formatting change) over the total number of edits during a period, represented as a percentage. Minor modifications require significantly less effort and time than other changes. *Ratio no comment* controls for the percentage of edits that have not been documented or classified by the editor. 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.

Time Invariant Control Variables

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 likely it is that someone will react to a new contribution by making edits.[25] *Age* of the article is measured as the log-transformed number of days that the article has existed prior to the last day in my dataset. The oldest articles tend to mirror topics in a traditional encyclopedia and to be longer than average since they have been available to edit for a longer period of time.

Another important set of article attributes that I control for are labels assigned by editors to articles, such as *categories*. On Wikipedia, categories represent narrow knowledge topics, such as “Mount Kilimanjaro” (Mountains of Tanzania) or “Albert Einstein” (Nobel laureates in Physics), or administrative categories.[26] The data I have acquired and used in the analysis aggregates 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.[27] 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 their registration.

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 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

Briefly summarized, the proposed mechanism states that increased demand for a good increases the likelihood of more consumer contributions, which in turn increase the number of expert contributions ultimately affecting the volume and quality of the good produced. The results detailed below provide strong support for these hypotheses and hence for the social mechanism aligning demand with collective production. The positive correlation between Wikipedia article consumption and quality (0.22) and, respectively, consumption and article length (0.35), indicate that, conditional on consumer participation and expert response, the quality and volume of needed goods are higher than those of goods which are less in demand.

Insert Tables 4-6 about here

Effects of demand on novice production. Table 4 presents four regression models based on equation (2) containing article consumption, with the number of novice edits to article k during period t as a dependent variable. As predicted by the first hypothesis, one unit increase in demand for an article, measured as log-transformed article views, results in an increased likelihood of novice contributions during the consumption interval. This finding is consistent with my qualitative research findings indicating that novices contribute to articles mainly as a consequence of consuming them. Models 1 and 2 present the results of the regressions using fixed-effects negative binomial estimators. Consistent with Hypothesis 1, I find that the more views the article has, the more likely novices are to edit it. A one-point increase in log-transformed article views translates into an increase of 21% in the number of novice contributions. Models 3 and 4 show results from a random-effects negative binominal analysis using the same specifications as model 1 and, respectively, model 2, which allows me to test the robustness of fixed-effect models.[28] I find that the effect of consumption in random-effect negative binomial models with the same specification is statistically significant and positive, but lower (over 21% in models 1 and 2 versus a 4.6% increase in the number of novice contributions for an increase in one unit of log-transformed article views in models 3 and 4, according to the incidence rate ratio values).[29]

Effects of demand and novice production on expert production. I show tests for Hypotheses 2a and 2b in Table 5 using four different regression models based on equation (3) with fixed- and random-effect negative binomial estimators. The dependent variable in Table 5 is the number of expert editors' contributions to article k during time (t+1). All covariates are lagged by one time 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.[30] In models 5 and 6, only time-variant controls were employed, due to the use of fixed-effects regression estimators. Models 7 and 8 include additional, 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 the article category.

The results in Table 5 confirm Hypothesis 2a, indicating that the direct effect of article consumption on expert editing patterns is statistically insignificant. Hypothesis 2b is also strongly supported across models 5 to 8: 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. Results presented in this table confirm that expert producers do not react directly to article consumption, because they are unable to observe it. Expert producers are stimulated to respond to article consumption *only if consumers signal demand* for that particular good through their contributions as novice producers.

Effects of novice and expert production on article length. Table 6 models 9 through 11 reports the results obtained from negative binomial estimates of the effect that novice and expert contributions have on the volume of good production based on equation (4). These models confirm Hypotheses 3a and 3b: the more expert contributions a good receives, the higher the volume of the good produced, and the more contributions from novices that a good receives, the higher the volume of the good produced. 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 50%

increase in article length compared to a 14-20% increase in article length for one unit in log-transformed novice edits. This difference suggests that expert contributions have a significantly stronger impact on the volume of good produced than novice contributions, possibly because experts often contribute additional, substantial information from published data sources whereas many novices make only occasional copy-editing contributions.

Models 10 and 11 also indicate the effects of control variables on article length. I find that importance is positively correlated with length, such that an important article is 18-21% longer than a less important one. If we consider that more important articles are by definition more central to a knowledge domain, these articles are likely to be longer than less important or less sought-after articles. Consistent with the assumption that longer articles contain more information and that different people monitoring an article may be interested in different sections, I find that the more monitors an article has the longer it is likely to be. Articles that belong to more projects are also more likely to be longer; they probably contain information that pertains to more fields of knowledge. The number of undo's on the article is negatively correlated with page length in model 11, which is unsurprising given that undo's often represent deletion of material.[31]

Effects of novice and expert production on article quality. Table 6 models 12 through 14 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 based on equation (5). The results confirm that expert contributions have a statistically significant and positive effect on article quality (Hypothesis 4a), while novice contributions have a statistically significant and negative effect on article quality

(Hypothesis 4b). As expected, an increase in the number of expert contributions to the article increases the likelihood that the article is high-quality, whereas an increase in the number of novice contributions decreases this likelihood. These findings suggest 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.

The results confirm that the effect of article consumption on quality is statistically insignificant in all models (Hypothesis 4c), a finding consistent with my theory that goods in high demand are not consistently produced at a high level of quality. The addition of control variables in models 13 and 14 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.

Limitations and future work

This paper's findings provide evidence that collective production is better at satisfying demand when consumers express interest in certain heterogeneous collective goods through novice participation in production. I find that an increase in collective good consumption increases the likelihood that consumers will contribute to production, and that these contributions result in subsequent contributions by expert producers, while consumption itself does not directly affect expert producer contributions. Additionally, I show that both novices and experts can increase the volume of collective goods produced, but only experts increase the quality of the goods (novice contributions decrease quality). This latter finding suggests that the alignment of collective production with consumption is subject to a tradeoff between the need for consumer participation to signal unsatisfied demand to expert producers, and the danger that consumer participation may decrease the quality of the needed goods.

The results I present provide strong support for this theory, but they are not without shortcomings. First, I summarize the way I have addressed potential estimation biases arising from the structure of the data and the choice of methodology. Second, I address concerns regarding the variability in types of novice and expert contributions. Although a more precise measure may provide a more fine-grained insight into the nature of the theorized mechanism, I argue that measuring overall contribution frequencies increases the generalizability of this study by distancing it from the particularities of Wikipedia contribution types. Third, I address the limitations of considering expert contributions as direct responses to novice editing. Lastly, I discuss concerns regarding the generalizability of the theory given the use of Wikipedia as an instance of collective action.

Endogeneity bias. Although the richness of this longitudinal dataset enables strong casual inferences, the analysis of article-level data could be subject to estimation biases stemming from endogeneity if any of the independent variables are correlated with the unobserved error term. There are several possible sources of endogeneity (Wooldridge 2006): the omission of a relevant variable, measurement error, or simultaneity bias, if a dependent and an independent variable affect each other at the same time. This concern is addressed in three main ways. First, the analyses employ a wide variety of potentially relevant variables as controls to capture the heterogeneity in article participation patterns and other characteristics, such as the age of the article, prior participation by experts and novices, prior participation by first-time editors, and prior article protection history. Second, to mitigate for the fact that unobserved article-level factors not accounted for statistically may correlate with article views or with expert and novice contributions, I use both fixed- and random-effects estimators, as well as Poisson QML estimators, all of which yield similar results.

Wikipedia contribution types. In Wikipedia, as in any type of collective action, producers may contribute in different ways. For example, experts can specialize not only in certain knowledge domains but also in formatting text, classifying pages, removing malevolent contributions, copy-editing, etc. Similarly, novice editors may make different contributions, from correcting spelling or punctuation to adding a reference or making a more substantial contribution of article content. The type of edit a novice makes may affect expert actions: an expert who notices a spelling correction from a novice may interpret this as interest in the article and spend more time improving it, while a substantial contribution from a novice may indicate a specific manner in which the article can be improved. Conversely, the type of expert observing a novice contribution affects whether the contribution is successful at stimulating additional work from experts: for example, a copy-editing expert who observes a novice attempt to improve the clarity of a paragraph may step in and further improve it, while an expert in classifying pages may be less likely to respond. The data available for this study is not able to make a distinction between different contribution types, so I use a rough measure of expert and novice contributions—number of edits—in my estimates. An analysis disentangling the effect of different types of novice contributions on the type and number of experts' response-contributions is the focus of a study I am currently developing.

Expert response. This study does not measure directly the response of expert editors to novice contributions. Instead, this response is inferred by examining changes in experts' editing patterns on articles following novices' contributions. Therefore, it is possible that 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 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.

Generalizability. The results of my analysis provide substantial support for all hypotheses. However, one could argue that the theory is limited by particularities such as Wikipedia's online nature or its technology of collaboration, which allows contributors to see the participation history of a peer producer and to a particular article. I argue that Wikipedia is not different in kind, but only in degree, in the sense that information is more readily available and participation records more transparent to contributors than in other settings. Contributors can also monitor changes to the volume or quality of collectively produced goods in offline situations as well: for example, one might periodically assess the state of the collectively produced good. Similarly, the production and consumption of collective goods are more difficult to measure in the offline world than in the digital realm, because interactions with the goods cannot be easily tracked. In the last section I discuss the identified social mechanism in relation to other market alignment mechanisms proposed in past research, and summarize the implications and theoretical contributions of my work.

IMPLICATIONS AND CONCLUSION

This article is based on a simple puzzle: under what conditions do collective producers know what goods are needed when they receive no direct information about demand? Given the fact that market alignment mechanisms such as prices are missing in the context of collective production, I argue that sociologists should pay more attention to the social mechanisms through which demand for collective goods is met. Accepting that collective production is fueled by intrinsic producer motivations and social incentives from other producers, I propose the existence of a social mechanism connecting collective production with consumption. In short, I show how contributions to collective goods production by novice producers are made in the context of the novices' consumption, and that these contributions are pivotal in drawing expert attention to goods that are in high demand. It is my hope that this paper will stimulate further work in examining the role that novice actors play in relation to demand satisfaction.

This study adopts an "analytical sociology" (Coleman 1986; Hedström and Bearman 2009) framework by investigating how a micro-level mechanism linking consumers and producers is influenced by, and influences macro-structures (here, the market for the collective good). My findings expand our understanding of social mechanisms by showing that expert contributions to the production of collective goods are most useful when they represent a response to consumer demand. These findings are of equal interest to economic sociology, through their contributions to understanding alignment mechanisms, and to collective action research, through their suggestion of a possible undersupply in some forms of collective action when consumer participation in production is lacking. I elaborate on these contributions below.

Economic sociology. Economic sociology has examined how categories, status, and embeddedness affect economic production decisions. To date, analyses of markets have highlighted the penalties that result when category signaling to end consumers or to intermediaries (such as brokers) is unclear (Zuckerman 1999), and the benefits reaped by firms when high-status actors compete for consumers (Podolny 1993). However, little attention dedicated to settings where producers are largely unaware of consumer demand due to a lack of direct alignment mechanisms.

Scholars have proposed two classes of incentives which affect production of a certain good: price and status-based incentives. In economic and social exchanges, individuals trade their time, effort, or resources they possess in exchange for equivalent goods, money, or social goods such as status. To the extent that an actor is willing to pay more for a good or service, that good or service is more likely to be produced for exchange.[32] Status incentives are important for a wider range of social exchanges, including economic ones, because even actors motivated by financial gains make production decisions linked to status in their industry (Podolny 1993). In enterprises that are less profit-oriented and where the goal is to contribute to collective goods such as scientific knowledge, individuals are rewarded not only financially through research funds, but also through collegial recognition and reputation which accrues to them for their work (Merton 1968; Reskin 1977).

William Goode's (1978) work on prestige as a control system suggests that there are similarities between economic and prestige markets, but many differences as well. For example, prices are an objective, known quantity that the producer or seller can demand in exchange for the good. In contrast, one cannot demand prestige or status rewards from others, and cannot negotiate in advance how much prestige one will receive for the action.

Actors receive social rewards when they produce a good that is desirable but costly or otherwise difficult for others to produce, and they are more likely to contribute to collective good production when they are rewarded with status or prestige for their efforts (Willer 2009). Individuals also contribute to collective good production due to social and intrinsic motivations such as learning and enjoyment (Lakhani and Wolf 2005) or a need to express identity or beliefs (Klandermans 2004). Organizational members have been shown to contribute to collective goods to help ease the tasks of others (Grant and Berry 2011), and scientists may take on a research question because they are dissatisfied with previous, less expert research or with unsubstantiated claims in the public media.

In my work, however, I argue that irrespective of underlying producers' motivations, novices' contributions to collective good provision help better allocate expert producers' efforts. I show that consumer contribution to production acts as a signal to expert participants of unsatisfied demand for a good. This is a simple mechanism through which the involvement of a beneficiary in the production process stimulates production in a manner which does not entail prestige or monetary exchanges. Simply allowing or encouraging the production system to be open enables feedback between the producers and consumers of collective goods. I suggest that this mechanism applies to a wide variety of sociological phenomena where expert producers of heterogeneous collective goods need to make decisions about the volume, quality, or features of a good, provided that the goods are collectively produced and that experts can observe novice contributions, as is the case with many cultural or information goods. In contexts where novices are rarely involved, such as the production of certain civic goods in the United States, there may be a risk of producing high-quality collective goods with low societal benefit, while goods in higher demand fail to be produced (Crenson and Ginsberg 2002; Skocpol 1999).

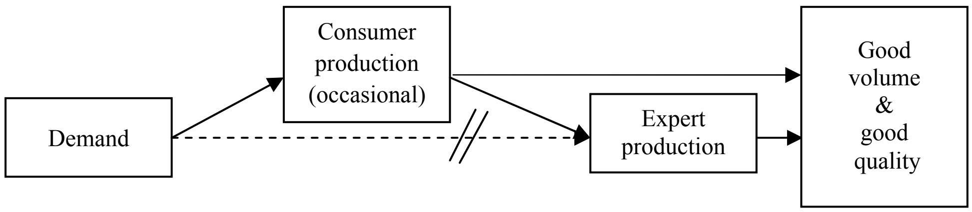
Collective action theory. Research on collective action has long focused on the central concern of free riding. Research on social movements has rarely asked how the demands and goals of social movements are crystallized and expressed, or what needs fail to be met when a social movement succeeds, although it has been acknowledged that what "comes to be [produced] as a collective good is the by-product of individual preferences and patterns of social relations" (Baldassarri 2009).

My study draws attention to the role played by novice participants in the collective production of heterogeneous, non-rival collective goods (Marwell and Oliver 1984; Oliver, Marwell, and Teixeira 1985). I show that when experts in collective action are unaware of consumer demand, or when consumers fail to voice their needs in a manner that is compelling to expert producers, demand may remain unmet. However, when novices attempt to contribute to production, they elicit a response from experts that leads to improved quality and volume of goods. This mechanism linking novices and experts is consistent with descriptions of the professionalization of social movements

(Staggenborg 1988), descriptions of collective action as characterized by a nucleus of highly-involved participants and a more diffuse network of supporters, and with observations that sociological research often originates in responses to popular (novice) beliefs (McGehee 1982). Examination of the process through which consumers' engagement in production acts as a sign of unmet demand for a good is a useful framework for advancing collective action research because it proposes that the success of collective action in producing goods may not result in demand being met, and shows that minimal consumer participation may be a solution for this deficiency.

Additionally, the results of this study highlight the existence of a tradeoff between consumer participation as novice producers in the collective production process, which is important as a sign of demand for goods, and the negative effect of consumer participation on the quality of the goods produced. This finding has implications for production systems that lack a good pricing mechanism, because it raises questions about whether these systems achieve benefits from consumer participation: that is, enough involvement to produce the things that are in demand, but not so much that quality is negatively affected.

Figure 1. Collective Production Mechanism



Notes: Peer expert rewards mechanism not represented in this figure.

Table 1. Wikipedia – Page Structure (Data used in this study comes from article history)

Page Type →	History	Discussion →	Discussion History
Article	Article History	Article Discussion	Discussion History
User Page	User Page History	User Discussion (Messages)	Discussion History
Policy /Infrastructure	Policy History	Policy Discussion	Discussion History

Table 2. Escalation in Norms and Policies on the English Wikipedia

GENERATION 1 (2001-2002)	GENERATION 2 (2002-2005)	GENERATION 3 (2005+)
Ignore All Rules (IAR)		
Neutral Point of View (NPOV)	NPOV	NPOV
Assume Good Faith (AGF)	AGF	
	Dispute resolution	Dispute resolution
	Semi-protection	Semi-protection
	Three Revert Rule (3RR)	3RR
	Criteria for Speedy Deletion (CSD)	CSD G1-12*
		CSD R1-3*
		CSD 11-8*
		CSD C,U,T,P*
		Biographies of Living People (BLP)
		Unsourced
		Nontagged
		Fair use images
		No spoiler tags
		Anon. article creation
		Reliable sources

Notes: Based on WikiSym 2010 keynote address by Andrew Lih. Each line represents a detailed policy or norm to be followed by Wikipedia participants in making their contributions.
*Third generation of CSD policy is invoked together with the reason for proposed deletion, which varies by type of page.

Table 3. Dataset Description

Datasets	Timespan	Level	Records
Article Edits	01/2001- 05/2009	Time-stamped edit	186,875,177
Article Views	10/2008 – 01/2009	Article, views/half month	8,339,418
Quality & Length	05/2010	Article	2,752,543
Monitoring	10/2009	Article	3,139,636
Category	10/2011	Article	2,002,826

Notes: Page Views data included other types of pages beside articles, such as disambiguations, redirects, and discussion pages.

Table 4. Negative Binomial Panel Estimates Predicting Novice Actors' Contributions to Article k during Interval t (**Hypothesis 1**)

	Novice contributions$_{k,t}$			
Independent	Fixed effects		Random effects	
Variables	(1)	(2)	(3)	(4)
Consumption$_{k,t}$	0.192***	0.194***	0.045***	0.045***
	(0.006)	(0.007)	(0.003)	(0.003)
Cumul. edits$_{k,t-1}$	-0.278***	-0.264***	0.269***	0.258***
	(0.015)	(0.015)	(0.006)	(0.006)
Protected$_{k,t}$		0.030**		-0.004
		(0.009)		(0.005)
Ratio no comment$_{k,t-1}$		-0.031		0.039
		(0.031)		(0.028)
Ratio minor$_{k,t-1}$		0.075**		-0.037
		(0.029)		(0.027)
No edits$_{k,t-1}$		0.072***		-0.131***
		(0.020)		(0.019)
Interval$_t$	No	Yes	No	Yes
Observations	93,332	93,332	168,739	168,739
Groups	11,880	11,880	21,986	21,986
Degrees of freedom	3	14	3	14
Log likelihood	38,388.43	38,249.89	71,900.93	71,712.99

Notes: Standard errors in parentheses. * p<0.05, ** p<0.01, *** p<.001 (two-tailed tests). Constant terms and a control for missing consumption (models 1-4) data were omitted from the table.

Table 5. Negative Binomial Panel Estimates Predicting Expert Actors' Edits on Article k during Interval t+1 **(Hythpotheses 2a** and **2b)**

	Expert contributions$_{k\ t+1}$			
Independent	Fixed effects		Random effects	
Variables	(5)	(6)	(7)	(8)
Novice edits$_{k,t}$ (log)	0.091***	0.061***	0.244***	0.185***
	(0.008)	(0.008)	(0.006)	(0.006)
Consumption$_{k,t}$ (C_k)	0.000	-0.002	0.002	0.004
	(0.003)	(0.003)	(0.003)	(0.003)
Cumul. novice edits$_{k,t}$	0.262***	0.192***	0.061***	0.044***
	(0.014)	(0.016)	(0.003)	(0.003)
Cumul. expert edits$_{k,t}$	-1.417***	-1.446***	0.088***	0.063***
	(0.022)	(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)
C_k* I_k (/10)			0.033***	0.024*
			(0.010)	(0.010)
C_k* P_k (/10)			0.049***	0.017
			(0.010)	(0.010)
C_k* M_k (/10)			0.071***	0.017*
			(0.008)	(0.008)
Category$_k$	No	No	Yes	Yes
Interval$_t$	No	Yes	No	Yes
Observations	86,480	86,480	144,847	144,847
Groups	10,932	10,932	21,246	21,246
Degrees of freedom	5	17	37	49
- Log likelihood	135,944.7	135,159.2	195,578.9	194,320.6

Notes: Standard errors in parentheses. * p<0.05, ** p<0.01, *** p<.001 (two-tailed tests). Constant term, controls for missing views, importance, and monitors, ratio no comments and first time edits omitted from the table.

Table 6. Negative Binomial Estimates Predicting the Volume (Length) and Logistic Regression Estimates Predicting the Quality of Article k **(Hypotheses 3a, 3b 4a, 4b** and **4c)**

Independent	Article length$_k$			Article quality$_k$		
Variables	Model 9	Model 10	Model 11	Model 12	Model 13	Model 14
Cumul. expert$_k$	0.409***	0.385***	0.394***	1.673***	1.565***	1.321***
	(0.018)	(0.020)	(0.026)	(0.093)	(0.099)	(0.143)
Cumul. novice$_k$	0.131***	0.124***	0.179***	-0.321***	-0.325***	-0.223*
	(0.018)	(0.019)	(0.017)	(0.067)	(0.071)	(0.096)
Avg.consum.$_k$ (AC$_k$)	0.014	0.015*	0.051***	-0.004	-0.002	0.079
	(0.008)	(0.008)	(0.013)	(0.028)	(0.028)	(0.067)
Length$_k$						0.587***
						(0.123)
Projects$_k$ (P$_k$)	0.035**	0.023*	0.077*	-0.097	-0.101	-0.135
	(0.011)	(0.011)	(0.032)	(0.053)	(0.052)	(0.146)
Importance$_k$ (I$_k$)		0.132***	0.190***		0.393***	0.502**
		(0.025)	(0.042)		(0.080)	(0.169)
Experts$_k$			0.007***			0.006
			(0.001)			(0.004)
Editors50%$_k$			-0.035***			-0.023
			(0.003)			(0.012)
Age$_k$			-0.239***			-0.694***
			(0.034)			(0.138)
Monitors$_k$(M$_k$)		0.074***	0.169***		0.139	0.366*
		(0.019)	(0.037)		(0.096)	(0.167)
Undo's$_k$		0.000	-0.004***		-0.003**	-0.007*
		(0.001)	(0.001)		(0.001)	(0.003)
Protected$_k$		-0.006	-0.016*		0.023	0.048
		(0.009)	(0.007)		(0.043)	(0.062)
AC$_k$* M$_k$ (/10)			-0.130**			-0.207
			(0.047)			(0.171)
AC$_k$* P$_k$ (/10)			-0.073			0.122
			(0.062)			(0.242)
AC$_k$* I$_k$ (/10)			-0.125*			-0.194
			(0.059)			(0.237)
Category$_k$	No	No	Yes	No	No	Yes
Observations	8,677	8,677	8,677	12,292	12,292	12,292
Deg. Freedom	5	11	38	5	11	36
-Log (pseudo) likelihood	80,519.01	80,416.10	52,308.43	1,205.70	1173.29	765.70

Notes: 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 about consumption, monitors, importance and article length were omitted from the table.

REFERENCES.

Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *The Quarterly Journal of Economics* 84:488-500.

Allison, Paul D. and Richard P. Waterman. 2002. "Fixed–Effects Negative Binomial Regression Models." *Sociological Methodology* 32:247-265.

Angrist, Jörn-Steffen and Joshua David Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Arrow, Kenneth J. 1977. "Toward a Theory of Price Adjustment." Pp. 380-90 in *Allocation of Economic Resources*, edited by M. Abramovitz. Palo Alto, CA: Stanford University Press.

Aspers, Patrik. 2009. "*How Are Markets Made?*" Working paper, Max-Planck-Institut für Gesellschaftsforschung.

Bacharach, Michael and Diego Gambetta. 2001. "Trust in Signs." Pp. 148-184 in *Trust in Society*, edited by K. Cook. New York: Russell Sage Foundation.

Baker, Wayne E. 1984. "The Social Structure of a National Securities Market." *American Journal of Sociology* 89:775-811.

Baldassarri, Delia. 2009. "Collective Action." Pp. 391-418 in *The Oxford Handbook of Analytical Sociology* edited by P. Hedström and P. Bearman. New York: Oxford University Press.

Beckert, Jens. 1996. "What Is Sociological about Economic Sociology? Uncertainty and the Embeddedness of Economic Action." *Theory and Society* 25:803-840.

Brown, John Seely and Paul Duguid. 1991. "Organizational Learning and Communities-of-Practice: Toward a Unified View of Learning, and Innovation." *Organization Science* 2:40-57.

Butler, Brian, Elizabeth Joyce, and Jacqueline Pike. 2008. "Don't Look Now, but We've Created a Bureaucracy: The Nature and Roles of Policies and Rules in Wikipedia." *Proceedings of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems*:1101-1110.

Coase, Ronald H. 1937. "The Nature of the Firm." *Economica* 4:386-405.

Coleman, James S. 1986. "Social Theory, Social Research, and a Theory of Action." *American Journal of Sociology* 91:1309-1335.

Coleman, James S. 1988. "Free Riders and Zealots: The Role of Social Networks." *Sociological Theory* 6:52-57.

Collins, Randall. 1981. "On the Microfoundations of Macrosociology." *The American Journal of Sociology* 86:984-1014.

Crenson, Matthew A. and Benjamin Ginsberg. 2002. *Downsizing Democracy*. Baltimore, MD: Hopkins University Press.

Eckstein, Otto and Gary Fromm. 1968. "The Price Equation." *The American Economic Review* 58:1159-1183.

Fernandez, Roberto M. and Doug McAdam. 1988. "Social Networks and Social Movements: Multiorganizational Fields and Recruitment to Mississippi Freedom Summer." *Sociological Forum* 3:357-382.

Fligstein, Neil. 2001. *The Architecture of Markets*. Princeton, NJ: Princeton University Press.

Franke, Nikolaus, Eric Von Hippel, and Martin Schreier. 2006. "Finding Commercially Attractive User Innovations: A Test of Lead User Theory." *Journal of Product Innovation Management* 23:301-315.

Giugni, Marco G. 1998. "Was It Worth the Effort? The Outcomes and Consequences of Social Movements." *Annual Review of Sociology* 24:371-393.

Goode, Willam J. 1978. *The Celebration of Heroes: Prestige as a Social Control System*. Berkeley, CA: University of California Press.

Granovetter, Mark. 1985. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology* 91:481-510.

Granovetter, Mark. 2005. "The Impact of Social Structure on Economic Outcomes." *The Journal of Economic Perspectives* 19:33-50.

Grant, Adam M. and Jim W. Berry. 2011. "The Necessity of Others is the Mother of Invention: Intrinsic and Prosocial Motivations, Perspective Taking, and Creativity." *Academy of Management Journal* 54:73-96.

Grant, Adam M. and Francesca Gino. 2010. "A Little Thanks Goes a Long Way: Explaining Why Gratitude Expressions Motivate Prosocial Behavior." *Journal Of Personality And Social Psychology* 98:946-955.

Greene, William H. 2000. *Econometric Analysis*. Upper Saddle River, NJ: Pearson-Prentice Hall.

Gulati, Ranjay. 1998. "Alliances and Networks." *Strategic Management Journal* 19:293-317.

Gulati, Ranjay and Martin Gargiulo. 1999. "Where Do Interorganizational Networks Come From?" *American Journal of Sociology* 104:1439-1493.

Hausman, Jerry A., Bronwyn H. Hall, and Zvi Griliches. 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." in *National Bureau of Economic Research, Cambridge, MA.*

Hausman, Jerry A. 1978. "Specification Tests in Econometrics." *Econometrica: Journal of the Econometric Society* 46:1251-1271.

Herstatt, Cornelius and Eric von Hippel. 1992. "From Experience: Developing New Product Concepts via the Lead User Method: A Case Study in a "Low-tech" Field." *Journal of Product Innovation Management* 9:213-221.

Klandermans, Bert. 2004. "The Demand and Supply of Participation: Social-psychological Correlates of Participation in Social Movements." Pp. 360-379 in *The Blackwell Companion to Social Movements*. New York City: Blackwell.

Kittur, Aniket, and Robert E. Kraut. 2008. "Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination." *CSCW Proceedings*.

Knight, Frank. 1971. *Risk, Uncertainty and Profit*. Chicago, IL: Chicago University Press.

Knoke, David. 1988. "Incentives in Collective Action Organizations." *American Sociological Review* 53:311-329.

Kollock, Peter. 1998. "Social dilemmas: The anatomy of cooperation." *Annual Review of Sociology* 24:183-214

Kyriacou, Andreas P. 2010. "Intrinsic Motivation and the Logic of Collective Action: The Impact of Selective Incentives." *American Journal of Economics and Sociology* 69:823-839.

Kriplean, Travis, Ivan Beschastnikh, and David W. McDonald. 2008. "Articulations of Wikiwork: Uncovering Valued Work in Wikipedia through Barnstars." *International Conference on Weblogs and Social Media* Proceedings:47-56.

Lakhani, Karim and Robert G. Wolf. 2005. "Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects." in

Perspectives on Free and Open Source Software, edited by J. Feller, B. Fitzgerald, S. Hissam, and K. Lakhani. Cambridge, MA: MIT Press.

Marwell, Gerald and Pamela Oliver. 1993. *The Critical Mass in Collective Action: A Micro-social Theory*. Cambridge, UK: Cambridge University Press.

Marwell, Gerald, Pamela Oliver, and Ralph Prahl. 1988. "Social Networks and Collective Action: A Theory of the Critical Mass. III." *American Journal of Sociology* 94:502-534.

McAdam, Doug. 1985. *Political Process and the Development of Black Insurgency: 1930-1970*. Chicago: University of Chicago Press.

McAdam, Doug, Sidney Tarrow, and Charles Tilly. 2001. *Dynamics of Contention*. Cambridge, UK: Cambridge University Press.

McGehee, Charles L. 1982. "Spiritual Values and Sociology: When We Have Debunked Everything, What Then." *The American Sociologist*:40-46.

Merton, Robert K. 1968. "The Mathhew Effect in Science. " *Science* 159(3810):56-63.

Okhuysen, Gerardo A. and Kathleen M. Eisenhardt. 2002. "Integrating Knowledge in Groups: How Formal Interventions Enable Flexibility." *Organization Science* 13:370-386.

Oliver, Pamela. 1984. "If You Don't Do it, Nobody Else Will: Active and Token Contributors to Local Collective Action." *American Sociological Review* 49:601-610.

Oliver, Pamela and Gerald Marwell. 1992. "Mobilizing Technologies for Collective Action." Pp. 251-272 in *Frontiers in Social Movement Theory*, edited by A. Morris and C. Mueller. New Haven, CT: Yale University Press.

Oliver, Pamela, Gerald Marwell, and Ruy Teixeira. 1985. "A Theory of the Critical Mass. I. Interdependence, Group Heterogeneity, and the Production of Collective Action." *American Journal of Sociology* 91:522-556.

Olson, Mancur. 1965. *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.

Ortega, Felipe. 2009. "Wikipedia: A Quantitative Analysis." Unpublished Ph.D. dissertation, University Rey Juan Carlos, Madrid, Spain.

Ostrom, Elinor, James Walker and Roy Gardner. 1992. "Covenants With and Without a Sword: Self-Governance is Possible." *American Political Science Review* 86(2): 404-417.

Ostrom, Elinor, Roy Gardner and James Walker. 1994. *Rules, Games, and Common-Pool Resources*. Ann Arbor, Michigan: University of Michigan Press.

Oz, Ayelet. 2008. "Move Along Now, Nothing to See Here: The Private Discussion Spheres of Wikipedia." *Berkman Working Paper*.

Passy, Florence and Marco G. Giugni. 2001. "Social Networks and Individual Perceptions: Explaining Differential Participation in Social Movements." *Sociological Forum* 16:123-153.

Podolny, Joel M. 1993. "A Status-Based Model of Market Competition." *American Journal of Sociology* 98:829-872.

Reskin, Barbara. 1977. "Scientific Productivity and the Reward Structure of Science." *American Sociological Review* 42(3):491-504.

Robson, Eleanor , M. Campbell-Kelly, M. Croarken, and R.G. Flood. 2003. "Introduction." Pp. 1-17 in *The History of Mathematical Tables from Sumer to Spreadsheets*, edited by M. Campbell-Kelly, M. Croarken, R. G. Flood, and E. Robson. Oxford, UK: Oxford University Press.

Schelling, Thomas. 1973. "Hockey Helmets, Concealed Weapons, and Daylight Saving." *Journal of Conflict Resolution* 17:381-428.

Shah, Sonali. 2000. "Sources and Patterns of Innovation in a Consumer Products Field: Innovations in Sporting Equipment." *Working paper, MIT*.

Silva, Santos J.M.C. and Silvana Tenreyro. 2006. "The Log of Gravity." *The Review of Economics and Statistics* 88:641-658.

Skocpol, Theda. 1999. "Advocates without Members: The Recent Transformation of American Civic Life." Pp. 461-509 in *Civic Engagement in American Democracy*. Washington, DC: Brookings Institute Press.

Spence, Michael. 1973. "Job Market Signaling." *The Quarterly Journal of Economics* 87:355-374.

Staggenborg, Suzanne. 1988. "The Consequences of Professionalization and Formalization in the Pro-choice Movement." *American Sociological Review* 53:585-605.

StataCorp. 2007. *Stata Statistical Software: Release 10*. College Station, TX: StataCorp LP.

Swedberg, Richard. 2005. "Markets in Society." Pp. 233-253 in *The Handbook of Economic Sociology*, edited by Neil J. Swedberg and Richard Smelser. Priceton, NJ: Princeton Univeristy Press.

Tarrow, Sidney. 1994. *Power in Movement: Social Movements, Collective Action and Politics*. New York: Cambridge University Press.

Thompson, James. 1967. *Organizations in Action: Social Science Bases of Administrative Theory*. New York: McGraw-Hill.

Urban, Glen L. and Eric von Hippel. 1988. "Lead User Analyses for the Development of New Industrial Products." *Management Science* 34:569-582.

Uzzi, Brian. 1996. "The Sources and Consequences of Embeddednes for the Economic Performance of Organizations." *American Sociological Review* 61:674-698.

Uzzi, Brian. 1999. "Embeddedness in the Making of Financial Capital: How Social Relations and Networks Benefit Firms Seeking Financing." *American Sociological Review* 64:481-505.

Weisbrod, Burton A. 1986. "Toward a Theory of the Voluntary Nonprofit Sector in a Three-Sector Economy." Pp. 21-44 in *The Economics of Nonprofit Institutions: Studies in Structure and Policy*, edited by S. Rose-Ackerman. New York: Oxford University Press.

White, Harrison C. 1981. "Where Do Markets Come from?" *Advances in Strategic Management* 17:323-350.

White, Harrison C. 2002. *Markets from Networks: Socioeconomic Models of Production*. Priceton, NJ: Princeton University Press.

Whitmeyer, Joseph M. 2007. "Prestige from the Provision of Collective Goods." *Social Forces* 85:1765-1786.

Wikipedia contributors. "Template:Grading scheme." Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/wiki/Template:Grading_scheme (last accessed September 14, 2011).

Wilkinson, Dennis. 2008. "Strong Regularities in Online Peer Production." *Proceedings of the 2008 ACM Conference on E-Commerce*:302-309.

Williamson, Oliver E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: The Free Press.

Willer, Rob. 2009. "Groups Reward Individual Sacrifice: The Status Solution to the Collective Action Problem." *American Sociological Review* 74:23.

Wilson, John. 2000. "Volunteering." *Annual Review of Sociology* 26:215-240.

Wooldridge, Jeffrey M. 2006. *Introductory Econometrics: A Modern Approach*. Mason, OH: South-Western.

Wu, Fang, Dennis M. Wilkinson, and Bernardo A Huberman. 2009. "Feedback Loops of Attention in Peer Production." *2009 International Conference on Computational Science and Engineering*.

Zhang, Xiaoquan and Feng Zhu. 2011. "Group Size and Incentive to Contribute: A Natural Experiment at Chinese Wikipedia." *American Economic Review* 101:1601-1615.

Zuckerman, Ezra. 1999. "The Categorical Imperative: Securities Analysts and Illegitimacy Discount." *American Journal of Sociology* 104:1398-1438.

ENDNOTES.

[1] One could argue that a market failure could occur not only because the market price fails to represent demand but also because marginal social benefit is lower than demand, or marginal social cost is higher than supply of goods. In such cases, supply might meet demand and lead to efficiency from a private perspective but not from a social perspective. In the case of common goods discussed here this is not a concern.

[2] This theory applies only to heterogeneous public good production where no actors can be excluded from accessing and using the goods, and actors may be interested in different parts of the good, such as political platforms or public garden features (Marwell & Oliver, 1993). For homogeneous public goods such as clean air or safe water, the needs of consumers and producers coincide by definition. For collective action resulting in club goods, the issue of mechanisms regulating the production and distribution of benefits has been examined in the literature on commons and common pool resources (Ostrom, Walker and Gardner 1992; Ostrom, Gardner and Walker 1994).

[3] Throughout this study, participant, producer, and contributor are used interchangeably to refer to individuals contributing to collective production activities.

[4] Even in situations where collective action contributors receive social rewards from consumers, such rewards are unlikely to be bestowed for producing a particular good rather than contributing to the group effort, as it is often difficult to differentiate specific contributions. In this situation, an individual who identifies with the collective production project may feel rewarded by consumers. This reward may alter how much effort an

individual puts into contributing to collective production, but not the choices he or she makes regarding which goods to produce.

[5] Good overproduction is not necessarily an issue as long as the goods that are necessary are produced as well. For example, artists or craftsmen who produce goods because of intrinsic interest in their creation may also cater to external audiences as a result of these choices. However, since resources such as time are limited and since multiple outcomes may compete for the same producers, as in the case of some social movements, (over-) production of one outcome may happen at the cost of investment in another.

[6] The signal is a by-product of novice attempts to improve on collectively produced goods and not an intentional, deliberate request for help from expert producers. It is irrelevant here whether novices are aware of this signaling effect, which is different than the meaning of "signaling" in other research (Bacharach and Gambetta 2001; Spence 1973).

[7] One could arguably aggregate demand by surveying a representative sample of the consumer population. However this survey would have to include a large number of possible goods, features or demands that consumers may be interested in, some of which may not even be known (a priori) to producers. In many cases, this wide-scale project may be prohibitively expensive and /or not feasible.

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

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

[10] 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. For more information, see Ortega (2009).

[11] The description of wiki software here is largely based on the software produced by the Wikimedia Foundation, the non-profit legal entity behind Wikipedia. Different types of wiki software vary in their feature set.

[12] The English Wikipedia had 11,405,052 pages by the time it had 2,183,496 articles (Ortega 2009). Less than a fifth of pages are articles because *Discussion, Editor* and rule pages are included in the page count.

[13] To select my interviewees I started with a theoretical sample of 50 registered contributors who had contributed between once 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.

[14] 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. 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.

[15] 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.

[16] As of mid-2009, after my data collection process had ended, Wikipedia introduced a feature that allows an editor on the Article History page to click through and examine pages views for the article. I do not have information on the usage of this feature but, based on informal conversations with expert Wikipedia contributors at WikiSym 2010 and Wikimania 2010, I believe that this feature has not altered the editing patterns of expert producers.

[17] 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).

[18] Only about 90% of these articles were started before October 1, 2008, such that we have data for all eight intervals; the rest of the articles have fewer than eight time intervals of data.

[19] I tested the sensitivity of the results to this restriction and found that coefficient estimates on the variables we use to test my hypotheses are still statistically significant in the expected direction for a cutoff of 50 edits.

[20] I acknowledge criticism levied (Allison and Waterman 2002) against the conditional negative binominal fixed-effects model proposed by Hausman et al. (1989) and implemented in STATA as xtnbreg (StataCorp 2007), according to which this is not a "true" fixed- effects method because it does not control for all time-invariant covariates. I do not, however, employ the estimator proposed by Allison and Waterman (2002) for an unconditional negative binomial model because their use of dummy variables to represent fixed effects raises estimation problems in a dataset such as my own with large N.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] Examples of administrative categories are templates (formatting standards for categories of articles), disambiguations (clarifications of different meanings for the same term), and redirects (which redirect readers from a page named after a less common term for a concept to the concept's main page, e.g., the "EU" page is a redirect to the "European Union" article). To the extent possible, templates, disambiguations and redirects as well as other types of administrative articles such as glossaries, lists, and images have been eliminated from the analyses.

[27] 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.

[28] The robustness of these results is also tested using a Poisson QML estimator; the coefficients on the demand variable are statistically significant but stronger than the fixed-effects estimator coefficients and are available upon request.

[29] This difference in estimation may originate from the fact that nearly 45% of articles in the dataset have no novice contributions, and are therefore not included in the fixed-effects estimation. This omission leads to a stronger relationship between demand and novice participation in the fixed-effects estimation compared to the random-effects estimation, which makes use of the full dataset.

[30] 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

[31] 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.

[32] Although production by consumers is not unique to collective production, I argue that this action has a different importance in economic settings. Researchers have found that "lead users" create markets by innovating and surfacing latent demand in areas not served

by commercial producers. This economic market dynamic differs from collective production by novices because (1) the observed lead users are highly knowledgeable participants and (2) commercial enterprises, in this case the “experts,” use lead user contributions to assess the existence of a profitable niche market (Franke, Von Hippel, and Schreier 2006; Herstatt and von Hippel 1992; Urban and von Hippel 1988).