Social Structure and Mechanisms of Collective Production: Evidence from Wikipedia

Dissertation Advisor: Mikołaj Jan Piskorski   Andreea Daniela Gorbatâi

SOCIAL STRUCTURE AND MECHANISMS OF COLLECTIVE PRODUCTION:

EVIDENCE FROM WIKIPEDIA

ABSTRACT

In my dissertation I propose three counterintuitive social mechanisms to alleviate the risk that collective production will fail to maintain participant involvement and respond to demand. My first study, based on a panel dataset of edits and views of articles in the English Wikipedia, shows that, although collective production lacks a price-like mechanism to estimate demand for the goods it produces, consumers' contributions act as such a signal to expert producers. In the second paper I examine the theory that collective production participation is greatest when social norms of collaboration are obeyed. Using a large panel dataset of production networks and norm-related behavior in Wikipedia, I show that social norm infringement is not completely detrimental to participation because norm enforcement increases the likelihood that the beneficiary producer continues participating. In my third paper, I rely on interviews with experienced Wikipedia producers to examine whether producers' ties to non-participants in collective production increase the likelihood of turnover, and whether producers' embeddedness in collective production reduces turnover risk. Surprisingly, I find that producers with networks rich in ties to non-producers and with a task-oriented approach to collective production are those least likely to stop participating.

TABLE OF CONTENTS

Page

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the mentorship, advice or support of many wonderful people in my academic and personal life. First, I want to thank my committee members, Misiek Piskorski, Peter Marsden, Richard Hackman and Toby Stuart. Misiek has been a wonderful mentor and a supportive coach throughout this entire project; I couldn't have asked for a more dedicated dissertation chair. Peter's faith in my academic abilities, and his support throughout my entire doctoral career, from my admission interview until the final feedback on my dissertation, has been unwavering and priceless. From the very first time I sat in his class, this work, and my academic career benefited from Richard's inquisitive nature, and his intellectual generosity and mentorship. He has inspired me to look at everyday settings with new eyes, refine my research questions and improve my analyses. And, last, but not least, Toby has provided great feedback on my quantitative analyses and on the theoretical framing of my work.

My thinking on collective production systems has been greatly aided by conversations with Nitin Nohria, Rakesh Khurana, Eric von Hippel, Karim Lakhani, Jason Beckfield, David Ager, Frank Barrett, participants in the Work, Organizations, and Markets seminar at Harvard Business School, and countless conference attendees and peers, among which I would like to thank Matthew Lee, Hila Lifshitz, Andrei Villarroel, Benjamin Mako-Hill, Andrea Forte, and James Howison. Other fellow students in the doctoral program, Anoop Menon, Sanjay Patnaik, Erin Reid, Sameer Srivastava, Ryan Buell, and Adam Kleinbaum have been a wellspring on intellectual inspiration and moral support throughout the completion of coursework, general exams, and dissertation draft deadlines. The advice and caring support of Janice McCormick provided invaluable help through the initial stages of my doctoral career at Harvard.

home; my long-time loves and close friends, Virgil Pavlu and David Kung for loving me, believing in me, inspiring me with their hard work and dedication and for having been my companions through far away travels and the (sometimes) drudgery of everyday life; my grandparents Mica, Bubu, Bunu and Buna, my dear cousin Raluca, and my parents, Mirela and Cristi for being in my life. Above all, I would like to thank D'Artagnan Horner. My loving partner, my husband, my love – thank you for bearing with me through the last couple of years of my Ph.D. and sharing my frustrations and joys, thank you for cleaning up the house without disturbing the endless piles of papers spread out on all of our tables (and chairs, and floor), despite how messy you think they are – and, most importantly, for your unwavering support and patience, for your warm and loving embrace, and for being a part of my family, my life and my future.

INTRODUCTION

Understanding the consequences of social structure is a problem of central concern for sociologists (Borgatti and Foster 2003; Emirbayer and Goodwin 1994). Research indicates that individual position in the social web of work and personal connections is highly consequential for individual effectiveness, gains, and even well-being. In my dissertation I examine the consequences of social structure for the success of collective production, where collective production is defined as a form of collective action in which a set of actors collaborate in producing collective goods or services. Collective production is a pervasive and consequential form of production in economic and social life. For example, social actors in a town may participate in the creation of collective goods such as a healthy community or a public park. Within organizations, actors may produce collective goods for the benefit of their peers such as a knowledge-sharing database or a communications network, a template to facilitate routine tasks or a recreational space. Organizations themselves may act collectively to produce changes to industry regulations and laws that affect other industry stakeholders. This dissertation proposes and tests three social mechanisms which alleviate the risk that collective production will fail to address demand for collective goods or retain contributors, and in doing so, tackles several puzzling assumptions and findings regarding markets and social networks.

The sociological literature on collective action has been largely centered on theoretical and experimental work (Chwe 1999; Gould 1993; Kollock 1998; Macy 1991; Marwell and Oliver 1984; Marwell, Oliver, and Prahl 1988; Oliver, Marwell, and Teixeira 1985; Pfaff and Kim 2003; Simpson and Macy 2004). So far, empirical research on collective action has focused on two main research questions regarding the success of collective production. The first question revolves around the success of collective production at attracting participants – asking who

participates, when, or why – and under what conditions. Research in this area either aims to solve the free rider problem directly or examines the social mechanisms through which actors join various forms of collective action, such as volunteer work or social movements. A second vein of research examines the success of collective action from the perspective of goal attainment, asking if the collective action succeeded in its purposes and whether the collective goods it aimed to create were produced.

I propose that even when these two conditions are met, collective production may still fail. The three risks of failure that I describe and address in my dissertation are the following: First, collective production may produce goods, but not those goods that are in most demand. Second, participants in collective production may exit as a result of suffering from norm infringement if their peer contributors are not respectful of collaboration norms. And lastly, participants may leave collective production if they are not satisfactorily embedded in their social networks, either because of inter-role conflicts between collective production participation and other roles – such as friend, employee or family member – or simply because their non-collective production social network may take priority over their collaborators in collective production.

In my first dissertation paper I propose a social mechanism that alleviates the risk that collective production fails to create goods that meet existing demand. Given that individuals who become involved in collective production are not a random sample of the population benefiting from collective goods (Wilson 2000) and that individual interest and social structure affect what goods are produced (Baldassarri 2009; Gronbjerg 1989; McAdam 1985; Shah 2000; Tarrow 1994), we would indeed expect that collective production is not well aligned with collective demand.

I begin by focusing on the puzzling fact that some goods in demand are created despite the absence of a price-like mechanism. We know that in economic markets price mechanisms serve to help align production and consumption of goods by signaling to producers how much consumers are willing to pay in exchange for a good or service (White 1981; White 2002). However, even if they were interested in producing goods that address existing demand, collective producers often have no direct information about what the beneficiaries of their work would like to consume. This is in contrast to economic producers who benefit from aggregate information in the form of prices (Hayek 2008) and a clear incentive to produce for material (exchange) purposes. Hence, the lack of information about existing demand could result in the production of goods that are not wanted by customers, while goods that consumers want are not produced. In my work, I examine the conditions under which collective production generates useful goods despite the absence of a price-like mechanism. I also show that consumer contributions play a critical role in signaling demand to producers. This leads to the counterintuitive conclusion that a production system that allows and even encourages potentially low-quality consumer contributions is better at addressing demand compared against production systems in which consumers do not – or cannot – contribute to production.

The second paper in my dissertation tackles the taken-for-granted sociological assumption that social network density (Granovetter 1985; Uehara 1990) plays a role in generating benefits for actors through reducing norm infringement and increasing social support and individual rewards (Coleman 1988a; Coleman 1988b; Coleman 1990). This study makes an important contribution because social norms that specify prohibited or prescribed actions are central to understanding social behavior (Macy 1991) and predicting the existence or breakdown of social order (Durkheim 1984). Different social science disciplines have proposed various

factors that can explain why people enforce social norms (Axelrod 1986; Bendor and Swistak 2001; Goode 1978).

Among the various factors which account for norm enforcement, sociologists have proposed social networks (Burt 1992; Simmel 1902). In his work James Coleman (1990) theorized a specific social mechanism through which social networks facilitate norm enforcement. He proposed that high network density structures offer more incentive for individuals to punish others for norm infringement because such actions are more likely to be observed and rewarded by third parties than in the case of sparser networks. In turn, Coleman reasoned, the knowledge that actors are more likely to punish norm infringers leads the latter to refrain from such activities, which reduces the incidence of norm infringements. This mechanism has been widely cited and used as a foundation for further sociological scholarship (Horne 2007). The correlation between network density and social norms has been demonstrated across a wide range of settings, from informal support groups (Uehara 1990) to banking (Uzzi 1999). However, these studies have not established the existence of a causal mechanism between social networks and norm-related behaviors, which raises questions regarding whether this relationship is sometimes spurious in nature (Elster 1989).

My research offers a test of this causal relationship by examining the relationship between network density and social norms in a longitudinal dataset, and by showing that the frequencies of norm infringement and enforcement vary with changes in social network density and that social norm enforcement elicits compensation from other actors. I use a longitudinal dataset from Wikipedia containing information about contributions to articles and participants' norm-related behaviors to empirically demonstrate the effects of social networks on reducing norm infringement and increasing norm enforcement. Although a priori, one would expect that

social norm infringement leads participants to exit a social network, I show that a moderate level of norm infringement actually benefits actors in the social network because this action gives participants the opportunity to bestow social rewards on each other through norm enforcement.

A third research paper focuses on the role that social networks within organizations and of those outside organizational life play in individual turnover decisions. One would expect that higher embeddedness and multiplexity of relationships that an individual has within an organization (such as both friendship and work-related ties) lead to lower turnover, and that multiple external roles and commitments are antecedents of turnover because multiple roles expand one's repertoire of behaviors and set of obligations – which may lead to conflicts (Goffman 1959). The idea of "role strain" and conflict is an important object of study in organizational behavior literature, particularly in regards to the work-family balance (Bielby and Bielby 1989; Greenhaus and Beutell 1985). Although it is frequently recognized that a person's work life needs to "be viewed in the context of family and personal concerns" (Kopelman, Greenhaus, and Connolly 1983), there is little research integrating these findings through the examination of social network participation at work or through the study of organizational exit decisions.

Predicated on the belief that individual turnover or continuance of collaborative production can be understood through one's narrative of participation, I use semi-structured interviews based on Atkinson's "life story" interview method (1998) to elicit information about interviewees' participation in the collective production of Wikipedia articles, with particular attention given to interpersonal dynamics. I then analyze the retrospective accounts of participation and triangulate them with objective participation information retrieved from archival data. My decision to adopt this methodological approach has been influenced by the

research question and by the desire to investigate previously unidentified mechanisms linking turnover and social network embeddedness within and outside organizations. I show that contributors whose involvement is oriented towards task completion instead of socialization with other producers, and whose other social networks (i.e., family, friendship and employment networks) are better developed are the contributors who are least likely to abandon collective production. I propose that high dependence on the social network of producers increases the risk of exiting the collective production as a result of power balancing operations (Emerson 1962). This unexpected finding addresses the increasing research interest regarding the impact of social networks outside organizations or industries on individual participation (Hardin 1982) and regarding the relationship between participant and non-participant social networks, and individual roles and identities (Hardin 1971; Marwell and Oliver 1984; Oliver 1993).

Overall this dissertation provides strong theoretical and empirical support for three social mechanisms that can alleviate the risk of collective production failures, hereby contributing to a number of research areas. First, despite excellent research work, especially in the area of social movements, the collective action literature has neither examined whether collective production is aligned with demand for collective goods nor proposed any demand revelation and satisfaction mechanisms. This work provides an answer to a previously unaddressed puzzle in this area: Under what circumstances is collective production aligned with existing demand for collective goods? The identification of the social mechanism through which participants can signal their demand for goods to producers and producers can observe the aggregate signal and respond represents a first step in understanding the strengths and weaknesses of collective production from the perspective of demand satisfaction. Future work in this area will examine the

circumstances under which this mechanism may fail to produce useful goods by studying the patterns of failure at each step in the identified mechanism.

Second, this work contributes to an existing stream of research on the importance of social networks for norm enforcement and restitution by offering quantitative evidence of a causal relationship between social networks and norms, and by highlighting the delicate equilibrium between the breakdown of social order due to norm infringement and the benefits of norm infringement as an opportunity to bestow rewards through enforcement, and also onto enforcing actors. Third, my work adds to the growing area of work / non-work research by proposing an application of power-dependence theory to the study of social mechanisms through which collective production participants who are deeply embedded may choose to abandon production. Based on Wikipedia data, the two latter contributions answer a scholarly demand for further investigation of minimal social capital and social structures that enable the functioning of technology-mediated social and economic transactions (Ibarra, Kilduff, and Tsai 2005).

Last, this dissertation builds on recent research (Burt 2011) that proposes online environments as a new research "context for familiar social processes." In my work I highlight the advantage of using an online collective production setting as a research site for studying the role that social structures of collective production play in production outcomes, and for identifying social mechanisms of production based on detailed information about individual actions and interactions, because such information is more difficult to capture in offline settings.

CHAPTER 1

ALIGNING COLLECTIVE PRODUCTION WITH DEMAND:

EVIDENCE FROM WIKIPEDIA

INTRODUCTION

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 1971). 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 offered by peers 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.[1] 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

---

[1] In this particular situation searches for an article, as reflected in Wikipedia's server logs, represent a good approximation of interest in the topic of the article, given that Wikipedia articles are easily accessible online, so interest can quickly and almost costlessly translate in search behavior.

contributors cannot observe article page views[2] —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

---

[2] 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.

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).[3] 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,

---

[3] 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.

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 1971;

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.[4] Making this assessment is challenging

---

[4] 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; many of these goods are often produced by governmental
organizations because they are of general interest. 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). Additionally,
this theory is concerned with the production of public goods by non-governmental organizations. While it may be
useful for governmental organizations to think whether they respond to citizens' demand for certain public goods,
government production of these goods is subjected to many other economic or political considerations that fall
outside the scope of this article.

because in the absence of aggregate information such as pricing it is difficult for producers[5] to know what consumers want to consume and whether their demands are satisfied.

A priori, there are a number of theoretical reasons to believe that there is significant misalignment between what is in demand and what is provided. Since it is difficult for producers to respond directly to consumer demand, they may produce goods that are not useful 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, status rewards and other social incentives (Coleman 1988b; Zhang and Zhu 2011) that promote collective production and make participants feel that they belong and are socially valued

---

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

(Grant and Gino 2010).[6] This logic suggests that collective production participants may be unaware of consumer wants and desires beyond their peer producers' and may fail to produce certain goods unless certain conditions are met.[7] 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), and 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

---

[6] 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.

[7] 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.

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 collective production process. I assume that the collective 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. Lastly, I assume that expert producers have a range of potential contributions they are capable and interested in making, but they are confronted with a scarcity of resources (time and effort dedicated to article writing) and do not have a stable set of preferences over which ones to produce first. One can think for example of an engineer building a small remote controlled plane: he may know how to make the internal mechanism of the plane but may not be aware of what particular features kids would enjoy. If he makes a plane for his two kids he could ask them what features to add, but if the makes a plane for their school trip, he may learn about features that kids like by observing the modifications that kids, while playing with the collective good he produced, have attempted to make to the plane, like adding a ribbon to the tail or particular colors on the wings.

The proposed mechanism works as follows. First, under certain conditions, consumers who seek insufficiently provided goods express their demand through an attempt to contribute as novice producers. When experts observe the novices' participation in the production of a good, they interpret novice participation as a sign of unsatisfied demand and respond by increasing contributions towards the creation of the good in question. In order for the alignment between collectively produced goods and demand to take place, experts need to be able to increase the

15

quality of goods in demand (see **Figure 1.1**). While it may be intuitive that, given the

opportunity, some consumers may attempt to participate in production and tautological that

expert production would improve the quality of a good, it is necessary to assert that both these

conditions have to be satisfied in order for existing demand to be met with a high-quality good. I

explain this mechanism in more detail in the next section.

```
┌──────────┐    ┌────────────┐    ┌────────────┐    ┌──────────┐
│          │    │  Consumer  │    │   Expert   │    │   Good   │
│  Demand  │───▶│ production │───▶│ production │───▶│ quality  │
│          │    │            │    │            │    │          │
└──────────┘    └────────────┘    └────────────┘    └──────────┘
```
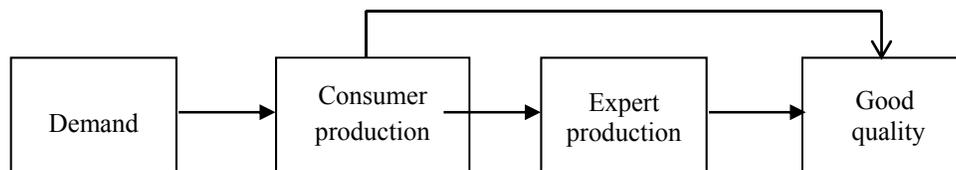
Figure 1.1. Collective Production Mechanism

*Notes***:** Expert production is also fueled by intrinsic motivation and by peer (expert co-participant) rewards, which are not represented in this figure and are outside the scope of this study.

Consumer demand and novice participation. When collectively produced goods are not

good enough to satisfy demand, consumers may choose one of several alternatives. They may

consume a substitute good or service, or continue using the collectively produced goods. In the

latter alternative, they may also contribute to collective production in an attempt to improve the

goods. 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, Campbell-Kelly, Croarken and

Flood 2003). A user of the R free software for statistical computing may attempt to use the R

Extension Package manual to create or improve, then share updates to missing or incomplete

statistical procedures. 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). 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 1971) and social

movement participation (Fernandez and McAdam 1988; Klandermans 2004; Passy and Giugni

2001).

However, assuming that potential consumers do not act to substitute the missing good for

another product, and that they are allowed to contribute to collective production, I propose that:

*Hypothesis 1***:** An increase in demand for a collective good is correlated with an increase

in consumer attempts to improve the good.

In other words, the more individuals seek to consume a good, the more likely it is that

some will have the knowledge and motivation to attempt to improve it. The propensity of

consumers to contribute may of course vary across goods with factors such as the initial quality

and the complexity of the good, the contribution cost (to the consumer) and also on the

consumer's willingness to use a substitute good, if any exist. However, deeply engaged

participants' contributions are driven by different mechanisms than the occasional, novice

contributions of consumers.

Novice participation as signaling to experts.[8] 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 only a few participants are deeply engaged and many are superficially

---

[8] The signal is a by-product of novice attempts to improve 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; this differs from the meaning of "signaling" in other research (Bacharach and Gambetta 2001; Spence 1973).

involved resonates with broader findings regarding reasons for involvement and non-involvement in collective action and social movements. First, "perceived effectiveness of one's own potential contribution" (Passy and Giugni 2001) and beliefs about the likelihood that others would participate (Oliver 1984) generate differential participation in collective action. Individuals who believe that their contributions are apt to make a difference, and those who do not expect that others will contribute to the production of useful goods are more likely to contribute to good production. Second, they may be motivated by selective incentives (Olson 1971) such as social rewards bestowed on them by their peers during the production process such as awards or deference.

Additionally, frequent contributors to collective production learn how to make effective contributions and coordinate with others to improve production processes and outcomes. 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).

When individuals dedicate substantial time and effort to furthering the goal of the collective production project, they become expert producers, 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. In many collective production settings, experts have a set of limited resources at their disposal such as time, information, and skills, and a set of independent and interdependent possible contributions. For example, experts working on a neighborhood park may have expertise in growing flowers, but the area does not have a garden in the plans. However, should

they decide to include a garden, flower experts have an array of possible flowers to plant and they may not know which of these flowers would be of more interest to the community benefiting from the park. Similarly, a social activist may be knowledgeable about social protection measures, but not be informed about the demand for social protection in a given community that is potentially affected by her social activism. Lastly, an IT department that creates a public good such as an intranet infrastructure for a large organization may have information about a wide range of designs, but not know precisely what features the large number of beneficiaries who are not involved in the production of the infrastructure would want to consume.

What I propose is that, while expert producers may not know what the beneficiaries of collective goods would consume, they may often observe if the latter attempt to contribute to collective production. For example, garden consumers may plant some flowers they would like to see, intranet users may attempt to install applications that fulfill a need not addressed by the IT producers. From such observations, experts who are deeply engaged in the production process may infer a latent, unsatisfied demand for a particular good or feature. As long as that good or feature is within their sphere of expertise and interest, experts may act to improve the good based on their own knowledge or by inferring from novice contributors what particular changes are desired by consumers. I predict therefore:

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

While expert producers may interpret novice producers' contributions as a signal of unsatisfied demand, in the case of many collective goods, it is virtually impossible for them to directly aggregate consumer demand in the same manner in which prices aggregate information

on markets.[9] For example, 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 or a park. Hence I propose:

> *Hypothesis 2b:* The effect of an increase in demand on contributions by expert producers is fully mediated by novice contributors.

Collective production outcomes. Increasing the volume of a collectively produced good is likely to be easier than increasing its quality. Contributing to collective goods by adding or removing features entails work that can be easily performed by novices and experts alike. 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. While it may be easy for both novices and experts to modify the goods, what is important for aligning collective production with demand is increasing the quality of the useful goods.

Compared against increasing good volume, 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

---

[9] 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.

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 the alignment of collective production to demand happens through this mechanism on condition that experts are capable of increasing good quality while novices are unable to directly improve good quality and may unintentionally lower it:

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

> *Hypothesis 3b***:** An increase in the number of contributions by novice producers will lead
> to reduced 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,[10] 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.[11] 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.

---

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

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

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.[12]

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 1.1**).[13] 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, authorship is obscured from consumers who read the final product. This lack of individual ownership and control is common in collective action phenomena.

Table 1.1. Wikipedia – Page Structure

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

Notes: The production data is this study comes from the article history.

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 policies and norms (see **Table 1.2** for the evolution of policies and norms in

---

[12] 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).

[13] 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.

Wikipedia).[14] 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).

Table 1.2. 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.

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).

---

[14] 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.

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 performing 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 sample of 35 expert contributors to Wikipedia.[15] Analyzing my expert

---

[15] 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.

producer interview data, I located 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;[16] 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.[17] 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

---

[16] 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.

[17] 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.

contribute only a few times (other novices contribute anonymously), and never become engaged

in the collaboration process. Given that repeated contribution leads to learning and socialization,

I assume a positive relationship between the number of contributions to Wikipedia articles and

expertise.

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. Another one explained:

"Because anyone can edit what you wrote, you have to go back and see if [your work] has been

vandalized, or if someone added some anything interesting to what you wrote. So when I come

back, I [may] add more [in response to the latter edits] and then come back to check it later."

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.[18] 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

---

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

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 don't agree [with them 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.[19] Some

experts explicitly collaborate: "I'm interested in are taxation issues, and there are a lot of us

interested in this topic. When you edit, and re-edit, each other's work, you start to build a

community, even if it's just a virtual one. I must say this community keeps me coming back."

However 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

---

[19] This process mirrors the learning through legitimate peripheral participation described by Lave and Wenger (1991).

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. 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, such that anyone with an Internet

connection can retrieve an article that she is interested in by performing a simple search, which

means the cost of expressing demand in the form of a page view is very low for consumers.

Third, encyclopedic articles are a type of good that requires minimal skills (i.e., literacy) to

consume, such that I can assume that an individual who views a page has "consumed" the

existing information.[20] 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

---

[20] For any type of collective good there may be individuals who do not benefit from it. For example, individuals who live in Chicago may not benefit from a public good like free air that is available to people in Boston unless they visit Boston. Since Wikipedia exists online, individuals without online access would not be able to consume this good, and some individuals consuming Wikipedia articles may be unable to assess their quality, or may be satisfied with a lower good quality than others (same as asthmatics may need a higher air quality than other individuals).This study proposes and identifies a social mechanism for alignment between demand and production of collective goods using online encyclopedic article production as a setting. It is therefore not concerned with issues such as online literacy, or digital divide or inequality in access to the internet and in internet skills which have been extensively discussed in the sociological literature (Carr 2007, Norris 2001, Schradie 2011).

them to take advantage of the full array of options for contributing and collaborating. Moreover, those who become experts are not a random sample of consumers, and this affects the likelihood that different topics are well-covered. One expert stated: "As with most tech things, my fellow editors tend to be web-savvy people from Western societies. Definitely more male than female, I would say by factor of 10 to 1. So the articles on developing countries are generally inferior to articles on developed countries, and there is a lot more stuff on Linux than there is on ballet, for instance. The arts in general are fairly under-represented compared to mechanical things, things that move." Another expert said: "I think contributors are mainly guys who are over-educated and under-employed. They have a lot of mental brainpower from all their years of schooling but are then doing really boring jobs. That I think is the core group of people who [work on] hundreds of articles a day." The non-representativeness of contributors is not unique to volunteer work Wikipedia; however, in this case, the digital traces of novice contributions are easier for researchers to identify than in other settings. Taken together, all these characteristics recommend Wikipedia as an 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 (see **Table 1.3**).[21] 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,[22] where an interval represents a half-month period for the production (article

edits) and consumption (article views) data between October 2008 and January 2009.

Table 1.3. 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.

As I have argued above, expert contributors on Wikipedia are those who have experience

collaborating 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

---

[21] 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).

[22] 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.

separate expert from novice contributors, where novices have a registered account or contribute anonymously.[23]

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 homoskedastic, 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:

$$L(\beta \mid y, X) = \prod_{k=1}^{N} \Pr(y_k \mid x_k) = \prod_{k=1}^{N} \frac{\Gamma(y + \alpha^{-1})}{y! \Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_k} \right)^{\alpha^{-1}} \left(\frac{\mu_k}{\alpha^{-1} + \mu_k} \right)^{y_k} \quad (1.1)$$

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

---

[23] 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. Note that expertise as defined here is time-varying in the sense that one may become an expert during the time interval analyzed, but more than half of registered participants edit less than three times. About one percent of the experts in the dataset became experts during each time interval analyzed.

heterogeneity is randomly distributed across articles, in order to use my time-invariant data.[24] 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 stable 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

---

[24] 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.

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):

$$E(novice - edits_{kt}|X_{kt}) = \exp(\varphi_0 + demand_{kt} * \varphi_1 + X_{kt} * \varphi + \theta_k + \varepsilon_{kt}) \qquad (1.2)$$

$$E(expert - edits_{kt+1}|X_{kt}) = \exp(\alpha_0 +$$

$$+ demand_{kt} * \alpha_1 + novice - edits_{kt} * \alpha_2 + X_{kt} * \alpha + \vartheta_k + \varepsilon_{kt+1}) \qquad (1.3)$$

$$E(expert - edits_{kt+1}|X_{kt}) =$$

$$= \exp(\gamma_0 + novice - edits_{kt} * \gamma_1 + X_{kt} * \gamma + \vartheta_k + \varepsilon_{kt+1}) \qquad (1.3')$$

where φ_0, φ_1, α_0, α_1, α_2, γ_0, and γ_1 are coefficients, and φ, α and γ 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 both by collective action theory about novice participation and by 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. I test Hypotheses 3a and 3b using a logistic regression model to evaluate the extent to which expert and novice contributions, 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,[25] 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 consumer 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.[26] The last equation takes the form

---

[25] 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. Since the number of topics an article relates to could affect the length and the quality of the article, I control for the number of WikiProjects an article belongs to in testing both Hypothesis 3 and 4.

[26] 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

$$\text{Quality} = \frac{exp(\gamma_0 + cumul.novice_k * \gamma_1 + cumul.expert_k * \gamma_2 + X_k * \gamma + \varepsilon_k)}{1 + exp(\gamma_0 + cumul.novice_k * \gamma_1 + cumul.expert_k * \gamma_2 + X_k * \gamma + \varepsilon_k)} \qquad (1.4)$$

where coefficients γ_0, γ_1, γ_2 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. Additionally, I test the relationship between article length at the end of the last interval and prior expert and novice contributions, both log-transformed. Length measured as the number of characters represents a reasonable metric of the volume of information a consumer receives on a particular subject, although it may be difficult to determine the comprehensiveness of the information based on length alone. This would ascertain that both experts and novices positively contribute to adding information to the articles, while only experts increase article quality. The cross-sectional negative binomial model employed in testing the effect of expert and novice contributions on article length is:

$$E(length_k | X_k) = exp(\beta_0 + demand_k * \beta_1 +$$

$$+ \ cumul.novice_k * \beta_2 + cumul.expert_k * \beta_3 + X_k * \beta + \varepsilon_k) \qquad (1.5)$$

where ε_k represents the error term, and β_0, β_1, β_2, and β_3 are coefficients and β vector of coefficients to be estimated.

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 and contributions of my research.

---

contestation by anyone concerned. Kittur and Kraut (2008) tested and confirmed the external validity of quality evaluations using ratings by non-Wikipedia participant readers.

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.[27] 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*).

---

[27] 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.

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.[28] 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 edits 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

---

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

likely it is that someone will react to a new contribution by making edits.[29] *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.[30] 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.[31] Articles labeled as important may

---

[29] 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.

[30] 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.

[31] 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

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 editors.

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 views and quality (0.22) and, respectively, views and article length (0.35), indicate that, conditional on consumer participation and expert response, the quality and volume of goods in demand are higher than those of goods that are less demanded.

Effects of demand on novice production. **Table 1.4** presents four regression models based on equation (2) containing article views, with the number of novice edits to article k during

---

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.

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 by a factor of 0.21 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-effects models.[32] I find that the effect of consumption in random-effects negative binomial models with the same specification is statistically significant and positive, but lower (by a factor of over 0.21 in models 1 and 2 versus by a factor of 0.46 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).[33]

While these results appear trivial, one could imagine situations where the number of novices contributing to an article does not increase when demand for the article increases. For example, a new quantum physics theory has been advanced, news of which may elicit higher

---

[32] 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.

[33] 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.

Table 1.4. Negative Binomial Panel Estimates Predicting Novice Actors' Edits to Article k during Interval t (**Hypothesis 1**)[34]

Independent Variables	Novice contributions$_{k,t}$			
	Fixed effects		Random effects	
	(1)	(2)	(3)	(4)
Views$_{k,t}$	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 views (models 1-4) data were omitted from the table.

demand for the article on the theory, yet many novices may not be able to contribute to the article describing it. Or, a school assignment to describe a geological phenomenon may elicit increased demand, in the form of 30 students looking up the article on Wikipedia, yet none may be able to make changes to the article. Also, the more demand for a good there is, the more consumers may expect that others (novices or experts) will contribute to it, leading to a free rider's dilemma (Olson 1971): such is for example the case for Mothers against Drunk Driving, an organization that clearly has appeal for a wide social group (mothers) yet not very high rates

---

[34] Key estimates in Table 1.4 and Table 1.5 vary across fixed and random effects models because the risk sets are different: articles with no novice contributions or, respectively, with no expert contributions (no variation in the dependent variable) are eliminated from the analysis in fixed-effects models, resulting in a smaller number of observations included in the analysis in the first two models of each table. These results are robust to a wide range of specifications.

of participation. At the extreme, a collective production process may be such that consumers are unable or unwilling to contribute. For example, in a context where deference to authority is high or the costs of publicly exposing oneself as a novice are high, consumers may not be willing to communicate demand by participating as novice contributors. Therefore, the condition that the more demand for a good the more novices are likely to contribute is necessary for the proposed alignment mechanism.

Effects of demand and novice production on expert production. I show tests for Hypotheses 2a and 2b in **Table 1.5** using four different regression models based on equation (3) with fixed- and random-effect negative binomial estimators and equation (3') with fixed-effects. The dependent variable in **Table 1.5** 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.[35] 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 time-invariant controls such as the number of projects the article belongs to, article importance, and the number of monitors; their interaction with article demand; and article category.

The results in **Table 1.5** confirm that **Hypothesis 2a** is strongly supported across models 6 to 8: edits by novice contributors have a statistically significant and positive effect on contributions by expert editors, ranging from an increase by a factor of 0.06 in fixed-effects negative binomial models to an increase of approximately 0.28 in random-effects models. **Hypothesis 2b**, stating that the direct effect of article consumption on expert editing patterns is

---

[35] 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

Table 1.5. Negative Binomial Panel Estimates Predicting Expert Actors' Edits on Article k during Interval t+1 (**Hypotheses 2a** and **2b**)

	Expert contributions$_{k\,t+1}$			
Independent	Fixed effects		Random effects	
Variables	(5)	(6)	(7)	(8)
Novice edits$_{k,t}$ (log)		0.061***	0.244***	0.185***
		(0.008)	(0.006)	(0.006)
Page views$_{k,t}$ ($V_{k,t}$)	0.146***	-0.002	0.002	0.004
	(0.007)	(0.003)	(0.003)	(0.003)
Cumul. novice edits$_{k,t}$	0.100***	0.192***	0.061***	0.044***
	(0.024)	(0.016)	(0.003)	(0.003)
Cumul. expert edits$_{k,t}$	-0.190***	-1.446***	0.088***	0.063***
	(0.031)	(0.024)	(0.004)	(0.004)
Protected$_{k,t}$		0.060***		0.004
		(0.011)		(0.003)
Ratio minor$_{k,t}$		-0.014		-0.057***
		(0.013)		(0.012)
No edits$_{k,t}$		0.083***		-0.050***
		(0.009)		(0.008)
Experts$_{k,t}$		0.002***		0.002***
		(0.000)		(0.000)
Importance$_k$ (I_k)			0.040***	0.029***
			(0.007)	(0.007)
Projects$_k$ (P_k)			0.003	0.004
			(0.003)	(0.003)
Monitors$_k$ (M_k)			0.032***	0.039***
			(0.006)	(0.005)
$V_{k,t} * I_k$ (/10)			0.033***	0.024*
			(0.010)	(0.010)
$V_{k,t} * P_k$ (/10)			0.049***	0.017
			(0.010)	(0.010)
$V_{k,t} * M_k$ (/10)			0.071***	0.017*
			(0.008)	(0.008)
Category$_k$	No	No	Yes	Yes
Interval$_t$	No	Yes	No	Yes
Observations	86,480	86,480	144,847	144,847
Groups	10,932	10,932	21,246	21,246
Degrees of freedom	4	17	37	49
- Log likelihood	136,028.1	135,159.2	195,578.9	194,320.6

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

fully mediated by novice contributions, is strongly supported in models 5 and 6. Results

presented in this table support the theory that experts are unaware of demand but they are

stimulated to respond to article consumption if consumers signal demand for that particular good

through their contributions as novice producers.

Effects of novice and expert production on article quality. **Table 1.6**, models 9 through

11, 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 (4). The

results confirm that expert contributions have a statistically significant and positive effect on

article quality (**Hypothesis 3a**), while novice contributions have a statistically significant and

negative effect on article quality (**Hypothesis 3b**). The addition of control variables in models 10

and 11 slightly reduces the positive effect of expert contributions and the negative effect of

novice contributions; however, the coefficients of interest remain statistically significant across

all models. Hence, an increase in the number of expert contributions to 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.

Table 1.6 models 12 through 14 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 (5). These models suggest that both expert and novice

contributions increase article length. There is, however, a difference in magnitude and the

disjunctive confidence intervals of the two coefficients: one additional unit in log-transformed

expert edits corresponds to a 50% increase in article length compared to a 14-20% increase in article length for one unit in log-transformed novice edits. This suggests that expert contributions have a significantly stronger impact on article length than novice contributions, possibly because experts often contribute additional, substantial information whereas many novices often make minor contributions.

Models 13 and 14 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 undos on the article is negatively correlated with page length in model 14 – this is unsurprising given that undos often represent deletion of material.[36] Overall, results in **Table 1.6** support the theory that while both experts and novices make positive contributions to article length, only expert contributions have a positive impact on article quality, while novice contributions decrease the quality of the article. In this situation, novice contribution may result in an expert response but not in an observable improvement of the collective good.

---

[36] 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.

Table 1.6. Logistic Regression Estimates Predicting the Quality of Article k and Negative Binomial Estimates Predicting Article k Length **(Hypotheses 3a and 3b)**

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

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 on average views, monitors, importance and article length were omitted from the table.

While the finding that experts increase article quality is hardly surprising, one could imagine situations where experts already made all possible contributions to the collective goods.

Another situation would be a case where the production of the collective good elicits both strong positive and negative externalities, such as for example controversial legislation like capital punishment, or controversial organizational policies that differentially affect social categories of employees. One can imagine that experts' advocating for either side, receiving input from "novices" regarding these issues may be unable to satisfy "demand" for these collective goods because of opposing interests by other experts. However given that the majority of articles in Wikipedia and, arguably, of other collective goods such as policies and legislation is not of such a controversial nature we would expect that experts are able to improve good quality.

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 most wanted goods.

The results I present provide strong support for this theory, but they are not without shortcomings. First, I explain how I have addressed potential estimation biases arising from the data structure 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. Last, I separate articles with strong demand shocks

49

(spikes in page views per period) from articles without demand shocks in order to examine whether the relation between novice and expert contributors could be observed because of demand shocks. I find that the results for the relationship between novice and expert contributions for articles without demand shocks are statistically significant and consistent with my theory.

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. Although I lack detailed level on contribution sizes, in future analyses I plan to distinguish between process experts and content experts by mapping expert activity on the relational matrix of articles.[37] Based on the data available for this study I am unable to distinguish among different contribution types, so I use a rough measure of expert and

---

[37] A preferable approach would be to look at article relationships from the perspective of co-searches. However for privacy purposes Wikipedia does not collect individual level search data. Therefore I plan to construct the relational article matrix in one of several ways: looking at inter-article links, looking at the overlap among knowledge categories across articles, or looking at editor co-participation patterns across articles. NLP algorithms may also be used to generate article similarity measures.

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 a positive, lagged effect of novice contributions on expert contributions identified in my quantitative analyses together with the reports of experts' reactions to novice contributions in my qualitative data suggest that the response to novice editing persists in the long run.

Additionally, there is a possibility that novices are more likely to contribute to certain articles. For example, interviews suggest that, given demand for an article, novices are more likely to make low cost contributions, either copyediting or contributing to articles they are familiar with, such as geography, or popular culture. Similarly, experts may be more likely to respond to novice contributions on certain articles rather than others. Further research is needed to understand how the patterns of novice contribution and expert response vary across article attributes.

In future work I am planning to test the existence of the proposed mechanism using data from a natural experiment on a smaller wiki-based site. The owners of the site randomly selected a set of articles, and they started emailing contributors information about the weekly demand for

articles they were involved in producing. Based on the proposed mechanism, I expect to observe a decrease or disappearance of the expert response to novice contributions when demand is revealed to experts, and statistically significant and positive expert response to novice contributions when demand remains obscured from experts.

Additionally there is opportunity to examine the extent to which the proposed mechanism applies to two other types of public goods. First, it is interesting to explore the extent to which experts are likely to respond to novice demand if the goods being produced are not collectively produced in the collaborative sense – that is, the novices do not alter the goods that the experts are working on. Do experts step in to respond to demand if the goods they are working are not altered? Second, one could study the extent to which experts respond to novice contributions when they are collaboratively working on producing innovative goods, as opposed to goods such as Wikipedia articles that are created through a synthesis of known facts about a knowledge domain.

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.[38] 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

---

[38] 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).

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 if they are dissatisfied with previous, less expert research or with unsubstantiated claims in the media.

In my work, however, I argue that irrespective of underlying producers' motivations, novices' contributions to collective good provision help to 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 (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 propose that when experts in collective action are unaware of consumer demand, or when consumers fail to voice their needs and desires in a manner that is compelling to expert producers, demand for collective goods may remain unmet. I then show that when novices attempt to contribute to production, they elicit a response from experts that leads to improvements in the quality and volume of goods in higher demand. 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.

In the following two chapters, I examine the social mechanisms through which the structure of collective production and the structure of social relationships beyond the collective production setting affect norm infringement and enforcement behavior, and, respectively, continued participation in collective production. When linking these to the present study that highlights the role of expert and novice contributions in collective production work, my overall research points to the importance of understanding the dynamics of collective production participation for the successful provision of collective goods.

TESTING COLEMAN'S SOCIAL-NORM ENFORCEMENT MECHANISM:

EVIDENCE FROM WIKIPEDIA[*]

INTRODUCTION

Sociologists have long invoked norms to explain social order (Durkheim 1984; Parsons 1953) and to account for various aspects of social behavior (Weber 1976). Norms embody a group's social consensus about appropriate behaviors. Some norms prohibit behaviors deemed unacceptable and specify punishments for flouting these proscriptions (Homans 1950:123). Others prescribe behaviors and reward those who undertake them (Blake and Davis 1964).

Among the various types of norms, sociologists have taken a particular interest in social norms. These norms require that parties personally unaffected by norm violation either punish offenders (Coleman 1990), or reward those who conform (Goode 1978). These characteristics of social norms raise a fundamental question: why one actor would punish or reward another for actions affecting others (Horne 2004). Since these rewards and punishments are costly for those who mete them out, but largely benefit others, potential enforcers are likely to have insufficient incentives to enforce norms (Olson 1971). In the absence of such enforcement, a second-order free rider problem develops and social norms are not observed (Coleman 1990; Oliver 1980).

Scholars from many disciplines have examined factors that lead people to enforce such social norms (e.g. Axelrod 1986; Bendor and Swistak 2001; Ellickson 2001; Fehr and Gächter 2002; Raub and Weesie 1990). Among these explanations, sociologists have particularly invoked

the role of network density (Burt 1982; Durkheim 1951; Lin 2001; Simmel 1902:170). Coleman (1990) formalized the argument, theorizing that high-density networks provide an opportunity structure within which third parties can compensate norm enforcers for the expense of chastising norm violators. Such payments encourage actors to punish those who violate norms, action which in turn reduces the incidence of norm violation.

Judging by the number of citations, Coleman's argument is now taken for granted in sociology (Horne 2001; Morgan and Sørensen 1999; Sampson, Raudenbush, and Earls 1997). There is also ample evidence of a negative correlation between network density and norm violations across numerous settings. Researchers have argued, for example, that norms against malfeasance among diamond traders and among geographically dispersed medieval Maghribi traders were sustained by high-density networks (Coleman 1990; Greif 1989). Similarly, in rotating-credit and informal help associations, the ability to sustain the norm of contributing to others' welfare has been shown to be associated with high density among associations' members (Barker 1993; Biggart 2001; Uehara 1990).

However, there is reason for skepticism that such correlational evidence can be used to support a causal link between network density and infrequent norm violations. First, some of the studies cited above examine a single social system and make inferences by pointing to the co-presence of network density and absence of norm violations without showing the counterfactual. Other studies that have undertaken comparative design were largely cross-sectional, making it difficult to establish causality. Furthermore, existing work provides little evidence to support Coleman's mechanism. This is problematic because simpler explanations can generate the same empirical predictions (Elster 2003). Consider, for example, a high-mutual-dependence environment in which actors exchange resources they value highly (Molm 1997). It is easy to see

that actors in such environments will violate norms infrequently, and that they will also establish dense relationships with one another (Horne 2001). In this case the relationship between density and norm violations is not causal but arises out of high mutual dependence (Flache and Macy 1996).

To provide evidence for Coleman's mechanism, it is necessary to follow his three-step reasoning process, and to provide support for each step using longitudinal data. Specifically, it is first necessary to confirm that norm violations decline as network density increases. Second, a researcher needs to furnish evidence that higher network density leads to more actions eliciting norm compliance (such as punishing norm violations), which then will lead to lower norm violations. The third step is to show that higher network density leads to more acts of compensating those who elicit norm compliance, which then leads to more acts of eliciting norm compliance, which then leads to fewer norm violations. Without supporting all three assertions, it is hard to assert that Coleman's argument has been tested properly. In this paper, we perform all three tests.

We undertake them in the context of the community of editors of Wikipedia, the largest user-generated on-line encyclopedia (Anthony, Smith, and Williamson 2009). This setting allows us to study norm violations in a naturalistic setting but at the same time to clearly observe (1) who violated a norm and who suffered from the violation, (2) who, if anyone, stepped in to punish, and (3) whether those who punished norm violators received rewards from the community for doing so. Also, because we observe actors over time as they experience transitions from a dense network to a sparse one (or vice-versa), we can provide results that are subject to fewer alternative interpretations. Finally, the network relationships we study are fairly weak, therefore providing very conservative tests of Coleman's theory.

The remainder of the paper is structured as follows. The next section examines the existing literature on norms, with particular emphasis on Coleman's mechanism, to derive our key hypotheses. We next describe our setting and data, and then our results. The final section discusses the limitations of our study and its conclusions.

THEORY

Step 1: Violating Norms

A norm is a set of rules specifying appropriate behaviors and backed by social rewards or sanctions (Blake and Davis 1964). Norms can be characterized on three dimensions. First, norms differ in their valence. Prescriptive norms encourage given actions, such as clapping at the end of a performance; proscriptive norms discourage specific actions, such as carrying a loaded gun. Second, norms differ in the types of behaviors they seek to regulate. Certain norms, often called *conventional*, seek to make everyone choose a single coordinated form of action that benefits all. Driving on the same side of the road is a conventional norm. Other types of norms resolve conflicts of interest between individuals and others. Often called *essential*, these norms mandate behavior that is beneficial to others but costly to the individual. Essential norms also prohibit behavior harmful to others but gratifying to the individual (Hechter 1987; Hechter and Kanazawa 1993). The norm not to pollute urban streets, for example, is beneficial to everyone but requires individuals to carry their trash rather than disposing of it on the spot.

It is easier to explain theoretically why actors comply with conventional norms than with essential norms. Because conventional norms are in everyone's interest, and no individual benefits arise from violating them, self-interested actors will comply with conventional norms. It is harder to understand why such actors comply with essential norms, since they bear the individual costs of compliance but appropriate only part of the benefit. This scenario leads to a

(first order) free-rider problem whereby every actor prefers not to comply with an essential norm but wants everyone else to do so. If all actors reason this way, no one will follow the norm. Thus, theoretical formulations of essential norms need to account for why self-interested actors comply with such norms.

Finally, norms also differ with regard to whether those who are expected to comply with them benefit from such behavior. Norms that benefit those who adhere to them are often called *conjoint*. A norm restricting use of a single telephone in a dormitory to ten minutes would fall into this category. At the other extreme are norms that do not benefit those who adhere to them, instead benefitting another group; such norms are usually called *disjoint*. An example is children who are expected to behave appropriately for the benefit of their adult caretakers. Most norms fall somewhere between the two ends of this spectrum, benefitting both those who comply with the norm and others who are not subject to it.

The distinction between conjoint and disjoint norms has implications for the free-rider problem associated with essential norms. In the case of conjoint norms, those who incur the cost of observing the norm are also its beneficiaries. Thus the free-rider problem is present but contained to a certain degree by the fact that individuals derive some of the benefits of their own normative behaviors. In the case of disjoint norms, those who incur the cost of following a norm are not its beneficiaries. Such absence of direct benefits accentuates the free-rider problem. This implies that it will be even more important for the beneficiary group to elicit norm compliance from the target group.

Step 2: Eliciting Norm Compliance

Given the difficulty of eliciting norm compliance, it is important to understand when and how it occurs (Bendor and Swistak 2001; Ellickson 1991; Homans 1950:123; Yamagishi and Cook

1993). In general, it is costly to elicit norm compliance. Resources used as rewards or punishments cannot be used for another purpose, but those who use their resources to elicit compliance enjoy only a fraction of its benefits. For most actors, the expected benefit will be too small relative to the cost; thus each actor will wait for others to elicit compliance. But if all potential actors behave in this way, no one will seek to elicit norm compliance. Coleman (1990) called this phenomenon "the second-order free-rider problem" to distinguish it from the first-order free-rider problem of compliance with norms described in Step 1.

The severity of the second-order free-rider problem depends on how actors seek to elicit norm compliance. Sometimes compliance is elicited with rewards. Goode (1978) argued, for example, that status can be used as a payment for complying with norms, particularly prescriptive norms. In other cases failure to comply with a norm elicits punishment, such as public chastisement. The distinction is important, because rewarding others rarely elicits negative reactions, whereas punishing others can easily prompt retaliation by those punished. As a consequence, actors are less likely to punish than to reward others (Molm 1997), and so the second-order free-rider problem is accentuated when punishment is used to elicit norm compliance (Horne 2007). It is thus particularly important to compensate those who punish others for failing to observe norms, a topic we will return to in Step 3 later.

Second, eliciting norm compliance can take the form of group or individual effort. When a group seeks to elicit norm compliance, each member can provide a small part of the reward or the punishment at a reasonably modest cost. Since the cost of eliciting norm compliance by a group is small, the second-order free-rider problem is attenuated (but not eliminated). In contrast, when a single individual is entirely responsible for eliciting norm compliance, he bears the entire

cost and the second-order free-rider problem is accentuated. Thus, compensating those who individually elicit norm compliance is particularly important.

Finally, norm compliance can be enforced either by those affected or by unaffected third parties. In most Western societies, for example, parents alone are expected to punish their misbehaving young children. Since those directly affected by the norm have a greater incentive to elicit compliance, the second-order free-rider problem is attenuated. For other norms, however, parties unaffected by a norm transgression are expected to step in and punish the offender. Norms backed by enforcement of this kind are often called *social norms*. For example, publicly disapproving of someone who fails to give up a seat on a bus for an elderly or a handicapped person is a social norm; unaffected third parties are expected to chastise someone who refuses to do so. Because such third parties bear the entire cost of eliciting norm compliance and appropriate none of the benefits, the second-order free-riding problem is quite strong. In such situations, compensating third parties for eliciting norm compliance is particularly important.

Step 3: Compensating Those Who Elicit Norm Compliance[39]

Despite the difficulties of eliciting norm compliance, it is possible to compensate those who engage in such acts. As before, compensating those who elicit norm compliance is subject to another free-rider problem, often called "the third-order free-rider problem" (Elster 2003; Horne 2001). The problem arises because such compensation is costly. Thus, each actor waits for others to provide compensation so as to appropriate the benefits without incurring the costs. This problem is most pronounced when compensation needs to be provided for punishing norm

---

[39] We use the word *elicit* to designate the act of inducing others to observe a norm, and *compensate* to designate the act of inducing others to elicit norm compliance. Thus, compensating logically precedes eliciting. Both actions can take the forms of giving rewards or administering punishments.

violators. As suggested above, eliciting norm compliance via punishment is more expensive than doing so via rewards, calling for a higher level of compensation.

A number of theories have sought to solve this third-order free-rider problem of compensation for punishing norm violations. Some theories invoked the intrinsic satisfaction derived from compensating others for eliciting norm compliance (Knutson 2004). Others suggested that bestowing rewards or punishments on those who compensate could solve the third-order free-rider problem. But this approach generates a fourth-order free-riding problem, leading to an infinite-regress problem (Elster 1989). To avoid this problem, most theories focused on solving the third-order free-rider problem directly. Specifically, a broad set of theories suggested that the third-order free-rider problem can be overcome when punishments for norm violations are compensated with rewards. This line of reasoning has gained substantial acceptance in the extensive evolutionary literature on norms, which shows that the rule of rewarding those who punish deviance wins in competition with other behavioral rules (Bendor and Swistak 2001; Opp 1982; Schotter 1982; Sugden 1986). In the same vein, most experimental results show that actors are most likely to observe norms when those who sanction norm violators are rewarded (Horne 2001). Finally, Coleman (1990) argued that compensation through rewards is less likely to suffer from the third-order free-riding problem, because rewards are cheaper to furnish than punishments.[40] Once this third-order free-riding problem is solved this way, Coleman argued that it is possible to solve both the second- and first-order problems and thus ensure that norms are observed.

---

[40] Coleman (1990:283) captured this argument stating: "Where sanctions are applied in support of a proscriptive norm and are consequently negative sanctions, the . . . problem of providing positive sanctions for the sanctioner is more easily overcome, because positive sanctions incur lower costs than do negative ones."

Coleman's Solution to the Free-Rider Problems

To understand Coleman's argument consider a numerical example with three actors, A, B and C. Assume that actor A considers whether to disobey a norm, which would bring personal benefits of $30 to A, but would also impose a cost of $30 on actor B and C each. This creates the first-order free rider problem, and will lead actor A to violate the norm unless he thinks he might be punished. Actor B or C could punish actor A for violating the norm, but assume that each would have to incur a cost of $35 to do so. If that's the case, neither actor B nor actor C will punish actor A. This leads to the second-order free rider problem. In principle, one of the affected actors, say C, could reward another, say B, for punishing actor A. However, actor B would have to receive a reward of, say, $40 to compensate him for the $35 cost to punish actor A. If actor C needs to incur substantial cost to provide this reward, say, also $40, he is he is unlikely to provide such a reward. He would rather suffer the $30 cost associated with norm violation than provide the $40 reward to B. The same logic applies to B rewarding C, leading to the third-order free rider problem.

Coleman argued, however, that there is a class of rewards that are very valuable to actor B, but cheap for actor C to furnish, as illustrated in **Figure 2.1** below.[41] Suppose that such rewards only cost $20 to actor C, but give $40 value to actor B. In this scenario, actor C will be willing to incur the cost of $20 to give such a reward, because he can avoid the cost of $30 when the norm is violated. This would solve the third-order free-riding problem. If actor B were to receive such a $40 reward, he would be happy to punish actor A because doing so only costs $35. This would solve the second-order free-riding problem. In anticipation of receiving the $35

---

[41] For simplicity of exposition, the figure only shows costs of norm violation borne by actor C and costs and rewards when actor C rewards actor B for punishing actor A. The scenario is symmetric for costs of norm violation borne by actor B.

punishment from actor B, it is no longer in Actor A's interest to violate the norm and obtain $30.

This solves the first-order free-riding problem and leads to the norm being observed.

Density and the Solution to the Third-Order Free-Rider Problem

Will not violate
norm

A

Cost of $35
if punished

Obtains $30
when norm violated

Cost of $35
to punish

Cost of $30
when norm violated

B

C

Incentive
to act

Value of $40
from reward

Costs $20
to reward

Incentive
to act
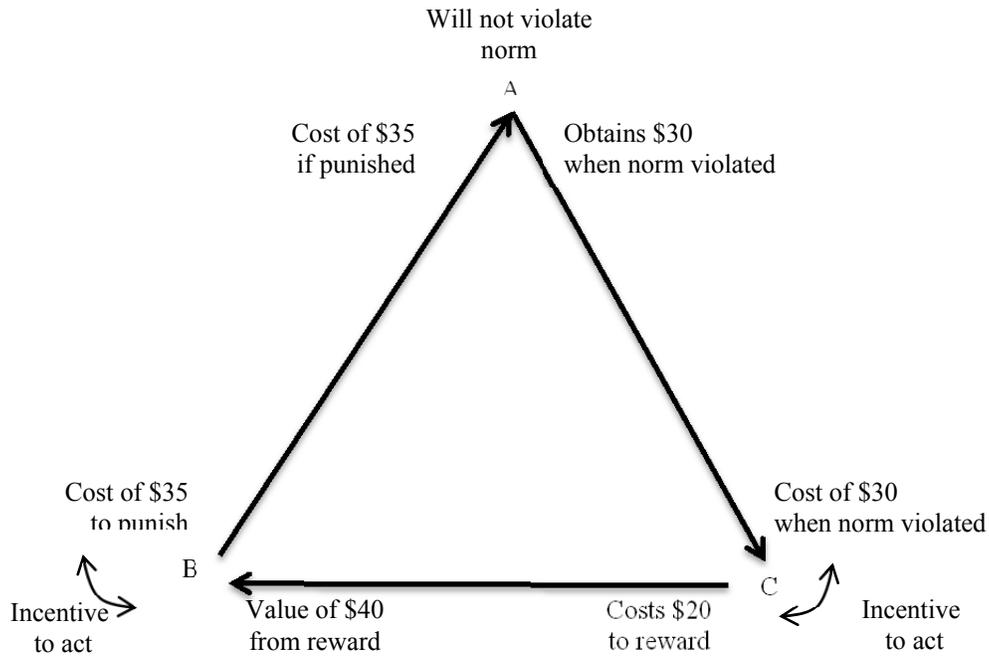
Figure 2.1. Solving the Third-Order Free-Riding Problem

Within this framework, it is easy to understand the role of network structures and in particular

the role of network density. Specifically, consider what would happen if there was no social

relationship between B and C, such that C could no longer reward B for the act of punishing A's

norm, as shown in **Figure 2.2**. Put simply, without C being able to compensate B, the second-

order free-rider problem cannot be solved. As a consequence, actor A will not be punished in the

event of norm violation, which will lead actor A to violate norms.

Will violate norm
A

Cost of $35
if punished

Obtains $30 when
norm violated

Cost of $30
when norm violated

Cost of $35
to punish

B

C

No incentive
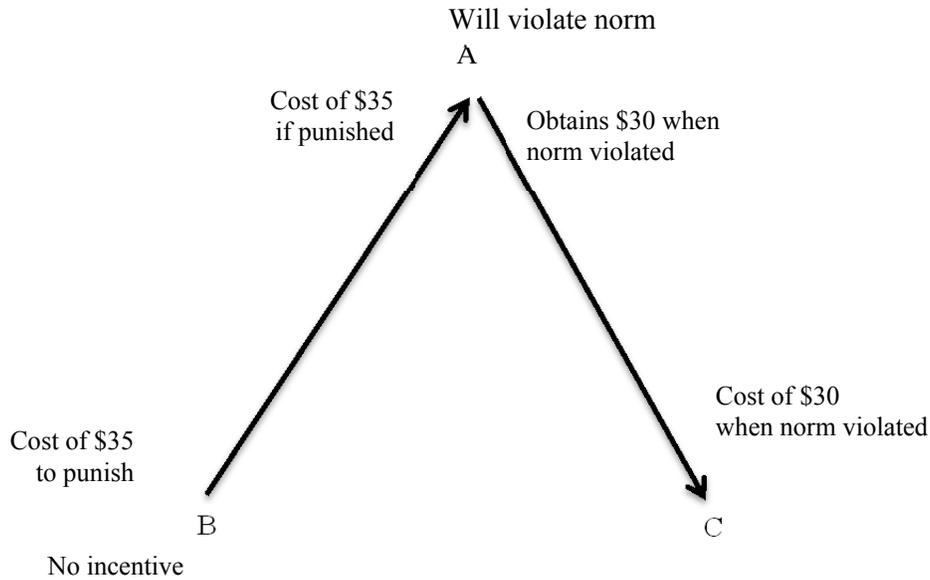
Figure 2.2. Failing to Solve the Third-Order Free-Riding Problem

This simple reasoning leads us to three pairs of hypotheses, one for each stage of the process. We will start with the final outcome of norm violations. Consistent with the discussion above, according to which actor A should engage in fewer norm violations in a high-density network, we argue that:

Hypothesis 1a: The higher an actor's network density, the less likely he or she is to

violate a norm.

Since actor A should violate norms less frequently when density is high, actor C should experience fewer norm violations too. Hence:

Hypothesis 1b: The higher an actor's network density, the less likely he or she is to

experience a norm violation.

For norm violations to occur less frequently under conditions of high density, it is necessary for those in dense network to punish norm infringements more frequently. In the example above, actor B administered such punishments in anticipation of rewards from C.

Because actor B is more likely to punish violations under conditions of high network density, we hypothesize:

> *Hypothesis 2a:* The higher an actor's network density, the more likely he or she is to punish a norm violation.

If in high-density networks actor B punishes actor A more frequently for inflicting norm violations against C, it should also be the case that actor C experiences more of punishments of actor A by actor B. Hence:

> *Hypothesis 2b:* The higher an actor's network density, the more likely others are to punish norm violators on his behalf.

Finally, for the entire mechanism to function, it should be the case that in high-density networks actor C rewards actor B more frequently for punishing actor A. Thus:

> *Hypothesis 3a:* The higher an actor's network density, the more likely he or she is to reward those who punish norm violations.

Since actor C is more likely to reward actor B under conditions of high density, it should also be the case that actor B obtains more rewards for punishing others under such conditions. Hence:

> *Hypothesis 3b:* The higher an actor's network density, the more likely he or she is to be rewarded for punishing a norm violation.

Commitment to a Social System

Thus far we have assumed that actors participate in a social system whether or not norm violations occur and whether or not violators are punished. In reality, however, actors can leave social systems to join others that will give them greater benefits. Such defections are particularly likely when actors are not heavily dependent on the social system and when they have easy

access to alternative social systems to meet their exchange needs. We assume that, in contemplating such a move, actors compare the utility they derive from the current system to the expected utility of joining another one. When actors experience norm violations, particularly violations that go unpunished, they experience negative utility. The benefits of staying in the current social system thus decline in comparison to the next best alternative, making actors more likely to leave. This reasoning leads us to the following hypothesis:

> *Hypothesis 4a:* An actor who experiences a norm violation whose perpetrator is not punished is less likely to continue participating in the social system.

Similar reasoning applies when an actor experiences a norm violation aimed at him or her, and chooses to punish the violator personally. Though the punisher may obtain some intrinsic benefits from doing so, he or she still incurs the costs of norm violation. As before, the benefits of staying in the current social system decline as compared to the next-best alternative, and he or she is more likely to leave the current social system. As a consequence, we hypothesize:

> *Hypothesis 4b:* An actor who experiences a targeted norm violation and personally punishes the violator is less likely to continue participating in the social system.

The same reasoning leads us to the opposite prediction when a norm violation is met with a third-party punishment. In this case, the target of the norm violation suffers its cost, but that cost is then offset by the benefit of seeing the offender punished without having to incur the cost of punishment. As a consequence, the target will end up at least as well off as if no norm violation had occurred. Furthermore, the target now knows that in this social system similar norm violations will meet with third-party punishments in the future. Consequently, when comparing the current social system to another social system in which norm violations may or

may not meet with punishments, he or she will be more likely to stay put. This reasoning leads us to hypothesize:

> *Hypothesis 4c:* An actor who experiences a targeted norm violation that is punished by a
>
> third party is more likely to continue participating in the social system than if no
>
> norm violation had occurred.

THE SETTING OF THE STUDY

We test our hypotheses in the context of contributions to Wikipedia, the largest on-line user-contributed encyclopedia. Between its launch in 2001 and the end of 2007, Wikipedia attracted over 6 million registered editors; these contributors created over 2 million encyclopedic articles in English and over 7 million entries in 253 languages. The site has become the seventh most visited website in the world.

Wikipedia was built on an intuitive on-line platform called wiki software. Anyone with internet access could post a draft of an article as long as the topic was deemed suitable for an encyclopedia. With the exception of a few protected articles, anyone could also edit any article by adding new content or by editing or deleting existing content. When an editor saved such changes, the software created a new version of the article for everyone to see. The previous version was added to the article history page, together with the Wikipedia username of the editor who had saved it and the time and date when the version was saved.[42]

No one could act as the final arbiter of an article's content; a subsequent editor could edit any version further. To manage disagreements over content, Wikipedia asked editors to try to

---

[42] Since Wikipedia did not require editors to register a personal account to make most types of edits, some editors made changes anonymously. In these cases, the Internet Protocol address of the computer where the edits originated was recorded. Many editors did open accounts, however; doing so allowed them to compile a record of their contributions, and provided personal pages where they could introduce themselves and receive feedback from other editors.

resolve differences of opinion via discussion. To ensure that such discussions did not interfere with article content, Wikipedia added a discussion page to each article. Wikipedia urged a focus on content and avoidance of *ad hominem* attacks and asked that editors act in good faith, signifying an intention to help the project rather than hurt it, and assume that others act in good faith in the absence of clear evidence to the contrary.

Wikipedia rules required articles to be written from a neutral point of view, which meant that they should fairly represent all significant views on the topic that had been published in reliable sources. It also required that article content represent and cite publicly available research. An editor's contribution to an article was considered acceptable as long as he could furnish reliable sources that readers or other editors could easily check.[43] Despite these rules and earnest efforts to reach consensus, editors could not always reach a viable compromise. At that point, participants could have recourse to a formal dispute-resolution process.[44]

Many editors were happy with Wikipedia's editing process. "*I don't have a problem with people making changes to what I wrote as long as they, you know, have good reasons for making those changes*," said one editor we interviewed. "*You know, like making the article better.*"[45]

---

[43] We use 'he' rather than 'he or she' because most Wikipedia contributors are men (see footnote 56 for details).

[44] The dispute-resolution process began with a "request for comment" from others, which allowed all editors to contribute their views on how the dispute should be resolved. Editors could also ask for assistance from a volunteer-run mediation committee or from volunteer Wikipedia editors who identified themselves as dispute-resolution specialists. If these measures proved insufficient, the mediation committee referred the case to the arbitration committee, staffed by 12–16 elected volunteers. That committee privately examined the entire record of all parties' conduct, paying particular attention to whether or not they had observed the good-faith rule. The committee then issued a public decision, which could ban an individual from engaging in particular behaviors or editing certain articles or from participating in Wikipedia in any fashion, either temporarily or permanently. The committee did not, however, rule on the "truth" of the underlying disagreement. By the end of 2006, the arbitration committee had ruled on over 100 cases.

[45] To collect interview data, we chose a random sample of editors from a list of current and past contributors available on Wikipedia. We contacted editors via e-mail and obtained a response rate of approximately 25 percent. We detected no response biases; the geographic and demographic profile of the editors we interviewed closely mirrors that of the entire Wikipedia population. At the request of editors, most interviews were undertaken via an instant-messaging program or free voice-over-IP programs. Interviews were analyzed using inductive methods to

Others found the process deeply troubling. "*There is no special treatment for experts or any way to bar anyone or group from changing the content*," said an editor who had stopped contributing. Indeed, Wikipedia's rules ensured that all editors were considered equal; no one's contributions were privileged by virtue of expertise in the field, advanced degrees or first-hand knowledge of the topic.[46]

Norm Violation: Undo

Wikipedia's open and democratic editing process made it possible for a kernel of an article to evolve very quickly into a full-fledged encyclopedia entry. It did, however, expose Wikipedia articles to acts of vandalism. Vandals—often unregistered editors—edited pages to add invective, deliberately replaced an entire article with invective or deleted article content altogether. To help editors recover valuable content after such acts of vandalism, Wikipedia attached an *undo* link to every version of the article on its history page (see **Figure 2.3**). By clicking that link, editors could swiftly undo the vandalized version of an article and replace it with the prior unaffected version. The vandalized version remained, however, in the history of article development. With this simple mechanism, Wikipedia editors were able to restore a page to its previous status as soon as an act of vandalism was detected.[47]

---

derive a theory of editor commitment, described in another paper by one of the authors. Quotes from the interviews are used here for illustrative purposes only.

[46] The only editors with special powers were a small group of administrators elected by consensus. These administrators were not employees of Wikimedia and did not enjoy special privileges when it came to content contributions or deciding on the value of others' contributions. They were, however, given the power to delete Wikipedia pages if the editor community voted to do so, and to block editors whose actions were deemed antisocial.

[47] Wikipedia also allowed registered editors to sign up for a watchlist on any page, which alerted them promptly to any changes on pages they were watching.

Figure 2.3.   Screenshot of a Sample Article History Page

The undo link could also be used incorrectly. Although its use was intended solely to undo vandalism, some editors found it an easy way to assert their points of view on article content. By clicking on the undo link, an editor could remove all changes introduced by the previous editor and restore the prior version without bothering to re-edit the content or negotiate with the other editor. Use of the undo link in the absence of vandalism constituted one of the biggest normative violations on Wikipedia. It flouted the basic tenets of acting in good faith and assuming that others do so as well. Many editors we interviewed also described the violation as such: "Imagine slogging over an article, trying to get all of the details right of something that happened 800 years ago, and then someone comes in and just erases you—no asking, no talking. . . . Poof, the content disappears! Can you imagine anything more disrespectful?" Indeed, many editors who had left Wikipedia cited instances of their work having been undone as a key reason for leaving. One former Wikipedia editor said: "I have a Ph.D. in South Asian musicology, so I really care that the Wikipedia entry reflects what we know about the topic. I spend a lot of time

documenting everything on the appropriate pages, and then . . . someone comes in and just undoes everything I have done. There is this one guy in particular does this all the time. So I try to talk some sense into him, but he won't talk. So I got really upset at all of this, and left."

The norm not to use the undo link (except when eradicating vandalism) has two characteristics that increase the likelihood that it will not be obeyed. First, as we described in Step 1 of the theory, this is an essential norm and therefore subject to first-order free-rider problems. All editors would prefer that the undo button not be used incorrectly and editors engage in the civil negotiation process over the article content. However, every one of them is tempted to use it to cheaply remove the content they disagree with. Second, this norm is at least partly disjoint in that those who are supposed to observe the norm (i.e. the editors of Wikipedia) are a smaller set than those who are the beneficiaries of norm compliance (e.g. the readers of Wikipedia who are not editors). This makes the free-rider problem ever more pronounced and again less likely that the norm will be observed. Given these conditions, any evidence of norm compliance should be seen as a conservative test of the underlying theory.

With these considerations in mind, we will treat use of the undo link as a norm violation (unless the undo removes profanity or reinstates an article after the bulk of its content has been removed). We will treat the editor who clicked the undo link as a norm violator. Consistent with **Hypothesis 1a** we expect that an editor embedded in a dense network will be less likely to undo an article version saved by someone else. Furthermore, our discussion indicates that all editors are affected by this violation, but the main victim of the violation is the editor whose version was undone. After all, he put the effort to contribute the content and it is his content that was removed. For this reason we will designate the editor whose version was undone as the main

victim of a norm violation. Consistent with **Hypothesis 1b** we expect that an editor embedded in a dense network will be less likely to experience an undo of an edit he saved.

Norm Punishment: Revert of Undo[48]

Because use of the undo link is readily apparent in an article's history, the author of an undone version and other editors will know that such a norm violation has occurred. Some editors ignore the undo and address the offending editor in good faith; others retaliate by clicking on the undo link themselves. This action, which undoes the previous undo and restores the prior version of the article, is known as *reverting the undo*. Because it conveys disrespect for the perpetrator of the first undo, that editor may respond with another undo, which may in turn be followed by another revert. Such skirmishes are known as "revert wars." To prevent them, Wikipedia has instituted a three-revert rule stipulating that no user can undertake more than three reverts on a given page within a twenty-four-hour period; violators are barred from making any changes to Wikipedia for a specified interval.

Many editors deal with undo actions on their own, but other editors and administrators can also step in to remind the offending editor that his or her actions are inappropriate. These reminders can take the form of a chastising note posted on the personal talk page of the editor in question; alternatively, a third-party editor can express disapproval more actively by reverting the undo. Like the original undo, which sends a public signal of disrespect, a revert of undo by an editor who is not the author of the undone version sends a public signal of condemnation of the undo act. It signals clearly that a third-party editor, uninvolved in the dispute, believes that

---

[48] On Wikipedia, the terms *revert* and *undo* are often used interchangeably. (See, for example, http://en.wikipedia.org/wiki/Help:Reverting#Undo). To prevent confusion, we will refer to the initial act as an *undo* and the act of undoing the undo as a *revert of undo*.

the original undo was unjustified and that its perpetrator violated a social norm and should be punished.

The punishment of an undo of a revert has three characteristics that make it less likely to occur. First, as we argued in Step 2 of the theory part of the paper, eliciting norm compliance through punishments rather than rewards makes it less likely to occur. Second, the act of a revert is individual rather than group effort. Again, as we argued in Step 2, this will make a punishment less likely to occur. Finally, using a revert of an undo gives us an opportunity to observe punishment of norm violation by an unaffected third party. As we suggested in Step 2 of the theory, such third-party punishments are particularly unlikely. Taken together, these three conditions imply that punishments through reverts are unlikely to occur, suggesting that we offer a conservative test of the theory.

With these considerations in mind, we will treat a revert of undo as a punishment of a norm violation. We will consider the editor who reverted the undo as the punisher, and the editor whose version was reverted as the punished actor.[49] Consistent with **Hypothesis 2a** we expect that an editor embedded in a dense network will be more likely to revert an undo of an article version saved by another editor. Consistent with **Hypothesis 2b** we expect that an editor embedded in a dense network will be more likely to experience other editors revert an undo of an article version saved by that editor.

Rewards for Punishing Norm Violators

Our interviews revealed that editors greet reverts with substantial gratitude. One commented:

---

[49] An editor other than the author of the undone version who undertakes a revert of undo may derive direct benefits from doing so if he or she cares about the quality of the article. If this is the case, the norm that the editor is enforcing is conjoint in nature. In the results part of this paper, we will distinguish between situations in which the reverter cares or does not care about the quality of the article and show that our results hold in both situations (see footnote 66).

People undo my work. It does not happen all that often, but more often than I would like. And then before I know it happened, someone will come to my rescue and revert the undo without even telling me. It's only later that I find what happened when I look at the article history. It's sometimes people that worked with me on that article . . . but you know what's most interesting? . . . It's also people who worked with me on other stuff . . . meaning they are kinda looking out for me! I would sometimes shoot them a note to say thank you. I would also definitely look out for them in the future to see if someone undoes their work and when that happens I would revert that… you know... as a way to say thank you for what they did for me.

This comment and others we collected along similar lines suggest that editors who revert undos receive rewards from victims of undos. Such rewards can take the form of written expressions of thanks or reciprocal reverts of undos. For the purposes of our paper we chose to use reciprocal reverts of undos as a measure of rewards for punishing norm violators.[50] We thus expect that editors who revert an undo will be rewarded in the future when third parties revert undos of their work. Specifically, consistent with **Hypothesis 3a** we expect that an editor is more likely to revert undos of article versions saved by other editors who themselves reverted other undos, and this effect is particularly large when the editor is embedded in a dense network. Furthermore, consistent with **Hypothesis 3b** we expect that an editor embedded in dense network will be more likely to experience other editors revert an undo of an article version saved by that editor if that editor has reverted other undos. We expect this effect to be particularly large when the editor is embedded in a dense network.[51]

---

[50] Alternatively, we could have used a measure of frequency with which editors are rewarded by obtaining a private or public thank you message. We chose not to use this measure, as it is very difficult to collect reliably.

[51] The use of reciprocal reverts of undos as a measure of rewards for punishing norm violators provides us with a conservative test of Coleman's mechanism. As we explained above, the mechanism works most powerfully when the reward for punishment is cheaper to supply than the punishment itself. In our case, rewards for punishment are captured by reciprocal reverts, and punishments are captured by reverts. Because the cost of undertaking a revert is similar to undertaking a reciprocal revert, Coleman's mechanism is likely to be weak. This makes it harder for us to detect evidence in support of that mechanism.

To test these hypotheses in the context of Wikipedia, we obtained a dataset from the Wikimedia

Foundation, the parent of Wikipedia, by downloading it from http://download.wikimedia.org.

The dataset contains every version of every article contributed to the English-language

Wikipedia site between January 2001 and October 2006.[52] For every article version, the dataset

provides the time and date it was saved, the Wikipedia username (or Internet Protocol address) of

the editor who saved it, and the version length in bytes. Having parsed the data, we wrote an

algorithm in MATLAB, described in **Appendix A**, to help us identify counter-normative undos

(i.e. excluding those that undid acts of vandalism) and reverts of undos.

Having run the algorithm across all articles in the dataset, we compared the resulting

statistics to Wikipedia statistics and to those reported in other papers that tried to identify acts of

undo and reverts of undo. The aggregate rates of undo and revert of undo identified by our

algorithm are very similar to those reported in related work – roughly 7% of edits are undos or

reverts of undos (Anthony, Smith, and Williamson 2009; Buriol, Castillo, Donato, Leonardi, and

Millozzi 2006; Kittur, Suh, Pendleton, and Chi 2007).

Dependent Variables

Armed with classifications of various sequences of article versions, we constructed the

dependent variables needed to test our hypotheses. We did so by aggregating the occurrence of

---

[52] The dataset also contains a complete record of discussion and talk pages, articles containing lists of other articles, and placeholder articles that merely redirect users to other pages. We exclude these auxiliary pages and analyze only the encyclopedia articles. The dataset does not include articles deleted prior to October 2006. This poses a potential problem, in that editors could have engaged in undo or revert actions on these articles, but a significant proportion of them were deleted because they contained very little content, and by implication generated little editing activity. Thus limiting ourselves to surviving articles does not substantially compromise our ability to detect acts of undo and revert of undo.

various norm violations and punishments over a month-long period (i.e. t = one month).[53] First, to capture the extent to which a given editor violated norms on Wikipedia, we constructed a variable *Number of Times Editor i Undid Others$_{it}$* equal to the number of instances when editor i undid any article version during time t.[54] We use this dependent variable in tests of **Hypothesis 1a**. To capture the extent to which an editor experienced norm violations, we constructed *Number of Times Editor i Was Undone$_{it}$* equal to the number of instances when editor i's article edits were undone by other editors during time t.[55] We use this dependent variable in tests of **Hypothesis 1b**.

To capture the extent to which a given editor punished others for violating the undo norm, we constructed two variables. To test **Hypothesis 2a**, we constructed a variable *Number of Times Editor i Reverted Others$_{it}$* equal to the number of instances during time t when editor i reverted an undo of a version that another editor j had saved. To test **Hypothesis 3a**, we constructed a variable *Number of Times Editor i Reverted Others Who Reverted$_{it}$* equal to the number of instances during time t when editor i reverted an undo of a version that another editor j had saved, as long as editor j had previously reverted another undo during time period t.

To capture the extent to which an editor i experienced others' punishing norm violations, we constructed three independent variables: (1) *Number of Times Editor i Was Undone Followed by No Revert$_{it}$* equal to the number of instances during time t when editor i's article edits were

---

[53] We also constructed the variables in two-week intervals, which increased the number of observations. Analyses using these variables produced equivalent results for the density measure across the models and yielded higher statistical significance. We thus report the more conservative results.

[54] We also constructed this measure using (1) the total number of edits, rather than instances, that were undone by editor i, and (2) an indicator variable that took the value of 1 if *Number of Times Editor i Undid Others$_{it}$* was greater than 0, and zero otherwise. The results are not sensitive to how we calculated this measure.

[55] This count does not include acts of undo by self. We also constructed this measure using (1) the total number of edits, rather than instances, by editor i that were undone, (2) an indicator variable that took the value of 1 if *Number of Times Editor i Was Undone$_{it}$* was greater than 0, and zero otherwise. The results are not sensitive to how we calculated this measure.

undone and received no reverts, (2) *Number of Times Editor i Was Undone Followed by Editor i Reverts Undo$_{it}$* equal to the number of instances during time *t* when editor *i*'s article edits were undone by others and then reverted by the focal editor *i*, and (3) *Number of Times Editor i Was Undone Followed by Another Editor Reverts Undo$_{it}$* equal to the number of instances during time *t* when editor *i*'s article edits were undone and then reverted by other editors. We use these dependent variables to test **Hypotheses 2b** and **3b**.

Finally, to capture the extent to which editors continue to contribute content to Wikipedia, we constructed a variable *At Least One Edit$_{it}$* equal to 1 if editor *i* undertook at least one edit (excluding acts of undoing others' edits) during time *t* and zero otherwise. We use these dependent variables to test **Hypotheses 4a**, **4b** and **4c**.

Independent Variables

Having defined our operationalization of the dependent variable, we now turn to the network of interactions among Wikipedia editors.[56] Editors rarely interact face-to-face, and most of their on-line interactions focus on content, rather than on socializing. As one editor said: "I'm probably one of the editors who is more prone than others to behaviors such as engaging people and getting consensus for difficult changes that people are struggling over. . . . But even then, I keep my engagement focused on factual contributions and not really on on-line socializing." By working together, however, editors developed close social bonds. In the words of one editor: "Even though you interact with people through text, it does tend to build community between

---

[56] Editors were volunteers, not employees of Wikipedia, and they did not receive direct monetary compensation for their contributions. According to Wikipedia's own surveys, over 86 percent identified themselves as male, and 70 percent reported being single. One-quarter were under 18 years old, one-quarter were between 18 and 22, one-quarter were between 23 and 30 and the remaining 25 percent were between 31 and 85. About one-third named a high-school diploma as their highest degree; 30 percent had an undergraduate degree and less than 20 percent had a master's degree or Ph.D. The same survey revealed wide variation in editors' motivations to contribute. "I liked the idea of sharing knowledge and want to contribute to it" and "I saw an error and wanted to fix it" were the two most frequently cited reasons for contributing. The least frequently cited reasons for contributing were a desire to make a reputation in the Wikipedia community, ambition to make money and fondness for mass collaboration.

editors. For example, I'm interested in taxation issues, and there are a lot of us interested in this topic. We have a really strong community, and I must say it keeps me coming back. If I was writing things completely in a vacuum, I would lose my interest." Not all editors were equally likely to experience such on-line relationships. One explained: "It's not like there is this one big Wikipedia community. There are communities inside the community. Some are strong; some are weaker. My personal experience is that most of the time, editing Wikipedia, I am doing it on my own and don't often encounter the same editors repeatedly."

On the basis of these statements we chose to use prior interactions between editors on the same articles as our measure of relationships between editors. To capture these interactions we wrote another algorithm which coded editor i and editor j as both contributing to the same article a if editor i had contributed at least one edit (excluding undos and reverts) to article a during period t, and editor j had contributed to the same article a during the same time period.

Some articles on Wikipedia, such as those about the World Cup, George W. Bush and Jesus, attract as many as 5,000 editors. It is hard to make the case that these editors interact with each other on these articles; many edit without being aware of each other's existence. By contrast, contributors to articles with fewer total editors are keenly aware of each other's existence and describe the process of editing as interaction. We thus decided to include only articles with fewer than 25 registered editors in our calculation of relationships between editor i and editor j.[57] We then used these data to construct a symmetric editor-to-editor matrix R_t, whose elements, r_{ijt}, consist of the number of articles with fewer than 25 total registered editors during

---

[57] We tested the sensitivity of our results to this restriction and found that coefficient estimates on the variables we use to test our hypotheses are still in the expected direction, though the statistical significance of the estimates is substantially lower across almost all of the specifications.

period t to which editors i and j both contributed during t.[58] On the basis of this matrix, we

constructed $\bar{r}_{ijt}$ equal to 1 if $r_{ijt} > 0$, and zero otherwise.

Using this definition we constructed a simple measure of density around editor i, we

calculate the number of relationships between editors with whom editor i has co-edited, as

represented by $\bar{r}_{ijt}$, and divide it by the number of possible relationships between editors with

whom editor i has co-edited. Using q as the total number of editors in the dataset at time t, we

defined this measure as: [59]

$$\text{Density}_{it} = \sum_{m=1}^{m=q}\left(\sum_{n=1}^{n=q} \bar{r}_{im} * \bar{r}_{mn}\right) \Big/ \left[\left(\sum_{m=1}^{m=q} \bar{r}_{im}\right) * \left(\sum_{m=1}^{m=q} \bar{r}_{im} - 1\right)\right] \qquad (2.1)$$

---

[58] It is also possible to define R_{ijt} as the number of total edits editor i contributed to articles to which editor j also contributed. This approach makes R_{ijt} asymmetric, and thus makes the empirical analysis more complicated. It also tends to make the relationship of i to j strong if i made numerous edits to a particular article. For this reason, we report the simpler analysis. Auxiliary analyses using the simpler approach yielded similar results but with weaker statistical significance.

[59] The results we report below are based on density measures using only the existence of a relationship, $\bar{r}_{ijt}$, rather than its strength r_{ijt}. In auxiliary analyses, we develop alternative measures using relationship strength and generate very similar results. We report results based on simpler variable definitions. We also test for one other specification of the density measure to protect ourselves from the following situation: three editors, i, j, and k, work on the same article; it is the only article they work on. If this is the case, i and j, j and k and k and i will each have a tie to each other and to no one else, and as such i, j, and k will be surrounded by a perfectly dense network. Such an environment would be likely to generate very few undos, and those that occurred would be quickly reverted by one of the three highly committed editors. As a consequence, we would observe a relationship among density, a low incidence of undos and a high incidence of reverts. This empirical observation would probably be an artifact of having three editors deeply committed to the article; it would have little to do with the mechanism we seek to test here. To protect ourselves from such a statistical artifact, we calculate another measure of density that excludes participation in the same article by j and k when i is present, given by $\tilde{r}_{ijkt}$. The formula is given by:

$$\text{Density}_{it} = \sum_{j=1}^{j=q}\left(\sum_{k=1}^{k=q}(\bar{r}_{ijt} * \bar{r}_{jkt} - \tilde{r}_{ijkt})\right) \Big/ \left[\left(\sum_{j=1}^{j=q} \bar{r}_{ijt}\right) * \left(\sum_{j=1}^{j=q} \bar{r}_{ijt} - 1\right)\right]$$

This is a very conservative estimate of triadic relationship between i, j, and k, because in this specification it is possible that editor i has co-edited with j on one article and with k on another article but is unaware that j and k co-edited another article together. If i is unaware of this relationship, he might fail to act in the manner described by the theory. Results for this specification are in the same direction as those for the variable defined in the body of the text, but the statistical significance of the results is often weaker.

Control Variables

Since density measures depend on the number of different articles editor i has edited, as well as the number of other editors who co-edited those articles, we include them as controls. First, we calculated a measure of *Number of Articles Edited$_{it}$*, equal to the number of articles that editor i edited during time t. Second, we captured the extent to which editor i edited the same articles repeatedly by constructing *Percentage of Articles Editor i Edited More than Twice$_{it}$*, equal to the number of articles with two or more edits by editor i during time t by the number of articles editor i edited during time t.

Third, we included *Network Size$_{it}$,* equal to the log of the total number of editors across all of the articles that editor i edited during time t. Fourth, we constructed variables *NetworkSize0$_{it}$* and *NetworkSize1$_{it}$* to reflect the fact that when variable *Network Size$_{it}$* takes the values of zero and one, it is impossible to define measures of density. In such situations, we assigned a value of zero to *Network Density$_{it}$*. To differentiate this zero from editors' actual scores of zero, we assigned a value of one to *Network Size0$_{it}$* when editor i edited other articles with no other editors during time t. Similarly, we assigned a value of one to *Network Size1$_{it}$* when editor i edited other articles with only one other editor during time t.

Finally, we included other measures that are not necessarily directly correlated with density but that can influence the extent to which editor i experiences norm violations or norm restitutions. First, we included *Cumulative Edits$_{it}$*, equal to the log of the cumulative number of edits by editor i prior to t, as well as the square of that number. Second, we constructed *Months since Signup$_{it}$* equal to the number of months since editor i first registered on Wikipedia. Finally, we constructed month dummy variables, *Time Period Dummies$_t$,* to control for temporal heterogeneity in norm violations, restitutions and project involvements.

Risk Set

It was our intention to examine undo and revert-of-undo actions by all registered editors in our dataset.[60] Our preliminary analyses revealed however that although there are over 600,000 editors in our dataset, almost 125,000 of them contributed only one edit, 50,000 edited only twice and roughly another 30,000 contributed no more than three edits. Our interviews revealed that such editors are unlikely to be familiar with Wikipedia rules, and are hence more likely to commit editing mistakes, e.g., to introduce a controversial point of view to the article without checking the article's talk page, where other editors may already have discussed how to handle this point of view. Edits by such inexperienced editors are often undone by existing editors without subsequent reverts of undo.

This dynamic is problematic for our analysis, because inexperienced editors have not had an opportunity to develop a dense network. Thus we are more likely to observe a positive relationship between low network density, a high incidence of undo and a low incidence of reverts of undo. Though this empirical observation is consistent with our predictions, it is not generated by the mechanism we want to test. To provide a more conservative test of our hypotheses, we chose to include an editor in the risk set only after he had contributed 25 edits, thus restricting our sample to 36,194 editors.

---

[60] It is possible to contribute to Wikipedia without registration, in which the edit is recorded together with the Internet Protocol address of the computer from which the change was made. Since it is possible that many different editors used the same computer to make changes (e.g. university library), we chose to exclude edits by unregistered editors and only focused on registered ones.

Table 2.1.   Descriptive Statistics

	Mean	Std. Dev.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 Network Density$_{i,t}$	.12	.26															
2 Network Size$_{i,t}$	2.79	1.94	-.15														
3 Network Size0$_{i,t}$	.20	.40	-.27	-.72													
4 Network Size1$_{i,t}$	.03	.17	-.10	-.19	-.09												
5 Number of Articles Edited$_{i,t}$	1.60	1.60	-.30	.45	-.59	-.13											
6 Percentage Articles Edited More than Twice$_{,t}$	.09	.19	.08	.15	-.24	.02	.13										
7 # Edits$_{i,t}$	1.70	1.69	-.21	-.61	-.32	-.09	.68	.12									
8 Cumulative Edits$_{i,t}$	3.62	2.12	-.28	-.30	-.27	-.12	.68	.07	.50								
9 Months since Signup$_{i,t}$	2.05	.95	-.06	.02	.08	-.03	.02	-.17	.01	.44							
10 # of Times Editor i Undid Others$_{i,t}$	.35	.80	-.16	.50	-.19	-.06	.56	.07	.71	.49	.07						
11 # of Times Editor i Was Undone$_{i,t}$	.41	2.67	-.06	.22	-.08	-.03	.26	.04	.24	.21	.02	.28					
12 # of Times Editor i Was Undone Followed by No Revert$_{i,t}$	.35	2.32	-.06	.21	-.07	-.03	.25	.03	.22	.20	.01	.25	.54				
13 # of Times Editor i Was Undone Followed by Editor i Reverts Undo$_{i,t}$	.04	.56	-.03	.11	-.04	-.01	.13	.04	.13	.12	.01	.21	.55	.36			
14 # of Times Editor i Was Undone Followed by Another Editor Reverts Undo$_{i,t}$	.06	.41	-.05	.17	-.06	-.02	.19	.03	.18	.15	.02	.22	.60	.48	.30		
15 # of Times Editor i Reverted Others$_{i,t}$	.23	2.66	-.02	.07	-.02	-.01	.07	.00	.08	.07	.02	.10	.03	.03	.01	.02	
16 # of Times Editor i Reverted Others Who Reverted$_{i,t}$	.14	1.74	-.02	.06	-.02	-.01	.07	.00	.07	.07	.02	.09	.03	.02	.01	.02	.58

Though this sample represents only a small subset of all Wikipedia editors, and thus raises issues of selection biases, it is reassuring to know that the editors in our sample contributed 70 percent of all edits.[61]

We then examined the timing of edits and found that as many as 40 percent of editors who edit one year do not do so the next year, suggesting that year-long data panels might be sufficient to capture most of the variation. We also found that 2005 was the most prolific complete year in our sample. That year alone witnessed the entry of more than 125 percent as many editors as there had been between 2001 and 2004. These editors contributed four times as many edits as they had between 2001 and 2004. Finally, in 2005, the rate of undos and reverts of undos increased more than twofold compared to the period between 2001 and 2004. For these reasons, we chose to focus on the time period between January and December 2005. **Table 2.2** provides descriptive statistics for the dataset analyzed.[62]

MODELS

To test **Hypotheses 1–3**, we used random-effects negative binomial models, which we constructed as follows. Following Hausman, Hall and Griliches (1984), we assumed that the value of the dependent variable for editor i at time t followed a Poisson distribution. On the basis of these assumptions, we construct a basic negative binomial model:

---

[61] In choosing the cutoff point, we considered the following tradeoff. An increase in the cutoff point reduced the number of editors in the sample and thus restricted the percentage of all edits under consideration. On the other hand, it increased editors' familiarity with Wikipedia's norms and made it less likely that we would be unable to define our density variables. (This consideration applied in particular to editors who edited articles singlehandedly without others' contributions.) We found that increasing the cutoff point to a minimum ofggg30 edits led to a substantial decrease in the percentage of all edits considered but had very little impact on our ability to define the density variable. On the other hand, lowering the cutoff point to a minimum of 20 edits had a much smaller effect on the percentage of total edits considered but a large effect on our ability to define the density variable. As a consequence, we chose 25 as our cutoff point.

[62] To test the robustness of our results, we re-ran our models for editors with more than 25 edits during 2004. The coefficient estimates on density variables remain in the predicted direction, but given the smaller frequency of undos and reverts, the statistical significance of the results is less robust.

$$\Pr\left(DepVar_{it} = depvar_{it} \,|\, x_{it}, \delta_i\right) = \left(\frac{1}{1+\delta_i}\right)^{\lambda_{it}} \left(\frac{\delta}{1+\delta_i}\right)^{depvar_{it}} \frac{\Gamma\left(\lambda_{it} + depvar_{it}\right)}{\Gamma\left(\lambda_{it}\right)\Gamma\left(depvar_{it} + 1\right)} \quad (2.2)$$

This model assumes, however, that the dispersion is constant across editors. To derive a random-effects negative binomial model, we allow δ_i to vary randomly across editors and assume $1/(1+\delta_i) \sim beta(r,s)$. Using f for the probability density function of δ_i we get the joint probability of the dependent variable for editor i at time t:[63]

$$\Pr\left(DepVar_{i1} = depvar_{i1}, \ldots, DepVar_{in} = depvar_{in} \,|\, x_{i1}, \ldots, x_{in}\right) =$$

$$= \int_0^\infty \prod_{i=1}^n \Pr\left(DepVar_{it} = depvar_{it} \,|\, x_{it}, \delta_i\right) f(\delta_i) d\delta_i =$$

$$= \frac{\Gamma(r+s)\Gamma\left(r + \sum_{t=1}^n \lambda_{it}\right)\Gamma\left(s + \sum_{t=1}^n depvar_{it}\right)}{\Gamma(r)\Gamma(s)\Gamma\left(r + s + \sum_{t=1}^n \lambda_{it} + \sum_{t=1}^n depvar_{it}\right)} \prod_{i=1}^n \frac{\Gamma\left(\lambda_{it} + depvar_{it}\right)}{\Gamma\left(\lambda_{it}\right)\Gamma\left(depvar_{it} + 1\right)} \quad (2.3)$$

RESULTS

Step 1: Violating Norms

With this specification, we estimated the likelihood that editor i violates norms on Wikipedia during time t by measuring the number of acts of undo undertaken by i against other editors, *DepVar$_{it}$ = Number of Times Editor i Undid Others$_{it}$*. It is possible that *Number of Times Editor i Undid Others$_{it}$* takes on a value of zero if editor i does not engage in any undo actions, whether absent from Wikipedia or fully engaged in the project. In order to focus only on situations when editor i performs no undo actions while fully engaged in Wikipedia, we want to control for

---

[63] We also constructed fixed effects negative binomial models as described by Allison and Waterman (2002) to remove all types of time invariant unobserved heterogeneity for editor i. Such estimations yield coefficient estimates on the main variables of interest that are directionally similar to those of random effects. However, the fixed effects estimation procedure assumes that the individual fixed effects is related to the individual dispersion parameter δ_i through a specific functional form, e.g. the fixed effect is the logarithm of the dispersion parameter (Hausman, Hall, and Griliches 1984). Guimaraes (2008) developed a method to test this assumption which we undertook on our data. We found that the test is not met, implying that the fixed effects negative binomial model might not perform reliably here. As a consequence, we prefer to report results from the more reliable random effects negative binomial, taking solace in the fact that the results are directionally similar across the two types of models.

situations in which he is absent from Wikipedia. We do so in a number of ways. First, we eliminate from the risk set for editor i all time periods t during which he did not contribute any edits. Second, we retain all time periods but include a dummy that takes the value of one when editor i contributed no edits during time t. Finally, we also use Heckman-like correction (Heckman 1979) and estimate the likelihood that editor i will contribute at least one edit during time t as a function of z_{it}, given by $Edits_{it} = z_{it}\psi + \omega_{it}$ where z_{it} includes selection-independent variables. Logit estimates of this equation are then used to derive the inverse Mills' ratio, given by $InverseMills_{it} = \phi(z_{it}\hat{\psi}) / \Phi(z_{it}\hat{\psi})$ where ϕ is probability density function, Φ is the cumulative normal density and $\hat{\psi}$ is the estimate of ψ. This ratio gives us the probability that editor i contributes at least one edit during time t given what we know about his or her characteristics as an editor. We include $InverseMills_{it}$ as an independent variable in the estimations of the *Number of Times Editor i Undid Others$_{it}$* model. All three methods yield the same results for the density measure, and for brevity we report only the uncorrected results and those with $InverseMills_{it}$.

Table 2.2 reports the results of these estimations. Consistent with **Hypothesis 1a**, we find that editors embedded in dense social networks are less likely to undo other editors' edits. This effect holds across all six models and is therefore robust to various specifications. As for control variables, we find that editors of articles that were not edited by anyone else were less likely to engage in acts of undo, and that those who co-edited with one other editor were no more likely to engage in such acts than those who co-edited with two editors. Beyond that, an increase in the number of co-editors led to an increase in the likelihood of performing an undo. Similarly, the total number of articles edited by editor i, as well as higher percentage of articles edited more than twice led to a higher incidence of engaging in an undo. Editors who had signed up a long time earlier were also more likely to perform undos, but that effect was offset by the negative

effect of actually contributing edits to a project. Finally, editors who had had their own edits undone by others were more likely to undo others' edits.

Table 2.2. Negative Binomial Random-Effects Estimates that Editor i Engaged in an Undo during Time t (Test of **Hypothesis 1a**)

Independent Variables	Number of Times Editor i Undid Others$_{it}$ Selection Equation: None			Number of Times Editor i Undid Others$_{it}$ Selection Equation: Editing		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Network Density$_{i,t-1}$	-.42**	-.42**	-.37**	-.32**	-.22**	-.21**
	(.03)	(.03)	(.03)	(.03)	(.03)	(.03)
Network Size$_{i,t-1}$	.24**	.24**	.21**	.24**	.23**	.24**
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
Network Size0$_{i,t-1}$	-.30**	-.26**	-.23**	-.08*	.08*	.07*
	(.03)	(.03)	(.03)	(.03)	(.03)	(.03)
Network Size1$_{i,t-1}$	.02	.02	.02	.09*	.16**	.14**
	(.05)	(.05)	(.05)	(.05)	(.05)	(.05)
Number of Articles Edited$_{i,t-1}$	.14**	.13**	.18**	.14**	.20**	.17**
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
Percentage of Articles Edited More than Twice$_{i,t-1}$	.65**	.63**	.62**	.55**	.61**	.59**
	(.03)	(.03)	(.03)	(.03)	(.03)	(.03)
Cumulative Edits$_{i,t-1}$	.07**	.08**	.04**	-.01	-.22**	-.14**
	(.00)	(.01)	(.01)	(.01)	(.01)	(.01)
Months since Signup$_{i,t-1}$	—	-.02*	-.07**	—	.39**	.42**
		(.01)	(.01)		(.02)	(.02)
Number of Times Editor i was Undone$_{i,t-1}$	—	-.01*	-.01**	—	-.01	-.01
		(.00)	(.00)		(.01)	(.01)
Number of Edits$_{i,t}$ x10	.05**	.07**	.06**	.07**	.12**	.08**
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
Inverse Mills Ratio$_{i,t}$	—	—	—	-.64**	-1.57**	-1.33**
				(.03)	(.06)	(.05)
Time Period Dummies$_t$	No	No	Yes	No	No	Yes
-Log-Likelihood	206,772	206,562	205,720	206,377	206,139	205,129
Degrees of Freedom	8	10	21	9	11	22
Wald χ^2	21,278	21,528	23,316	21,477	21,995	23,509

Note: Numbers in parentheses are standard errors. Constant was omitted. All χ^2 tests are based on a baseline model with no covariates. Results of sensitivity tests using generalized linear models with negative binomial link and grouped logits yield equivalent results. Selection equation predicting editing based on previous experience, tenure and month dummies was omitted from table. Resulting inverse Mills ratio was used as control in the outcome equation, as noted. Editors, i = 30,272; number of periods = 10; total number of observations = 212,317. (Not all editors started editing in time period 1.) *p < .05 **p < .01 ***p < .001 (two-tailed tests).

Table 2.3. Negative Binomial Random-Effects Estimates that Editor i Experienced an Undo during Time t (Test of **Hypothesis 1b**)

Independent Variables	Number of Times Editor i was Undone$_{it}$ Selection Equation: Edited			Number of Times Editor i was Undone$_{it}$ Selection Equation: Experienced undo		
	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
Network Density$_{it-1}$	-.18***	-.12**	-.27***	-.20***	-.25***	-.18***
	(.03)	(.04)	(.04)	(.03)	(.04)	(.04)
Network Size$_{it-1}$	.61***	.36***	.51***	.61***	.37***	.50***
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
Network Size0$_{it-1}$	—	-.32***	-.24***	—	-.58***	-.20***
	—	(.05)	(.05)	—	(.05)	(.05)
Network Size1$_{it-1}$	—	.06	.19***	—	-.03	.25***
	—	(.07)	(.07)	—	(.07)	(.07)
Number of Articles Edited$_{it-1}$	—	.23***	.24***	—	.20***	.21***
	—	(.01)	(.01)	—	(.01)	(.01)
Percentage of Articles Edited More than Twice$_{it-1}$	1.03***	.96***	.88***	1.07***	.97***	.83***
	(.03)	(.03)	(.03)	(.03)	(.03)	(.03)
Cumulative Edits$_{it-1}$	—	-.41***	-.36***	—	-.51***	-1.60***
	—	(.01)	(.02)	—	(.02)	(.05)
Months since Signup$_{it-1}$	—	.37***	.36***	—	.33***	1.37***
	—	(.02)	(.02)	—	(.02)	(.05)
Number of Times Editor i Undid Others$_{it-1}$	—	.42***	.44***	—	.40***	.44***
	—	(.01)	(.01)	—	(.01)	(.01)
Number of Edits$_{it}$ x10	.30***	.20***	.25***	.29***	.19***	.31***
	(.07)	(.09)	(.01)	(.07)	(.09)	(.01)
Inverse Mills Ratio$_{it}$	-.45***	-1.82***	-1.94***	-.24***	-1.51***	-5.90***
	(.03)	(.07)	(.09)	(.02)	(.08)	(.20)
Time Period Dummies$_t$	No	No	Yes	No	No	Yes
-Log-Likelihood	124,474	121,936	121,554	124,512	122,067	121,825
Degrees of Freedom	5	11	22	5	11	22
Wald χ^2	32,186	35,271	38,348	32,678	35,330	39,026

Note: Numbers in parentheses are standard errors. Constant was omitted. All χ^2 tests are based on a baseline model with no covariates. Selection equation predicting editing based on previous experience, tenure and month dummies was omitted from table. Resulting inverse Mills ratio was used as control in the outcome equation in Models 7,8, and 9 as noted. The second selection equation was used to predict whether editor i experienced an undo during time t with the same independent variables. Resulting inverse Mills ratio was used as control in the outcome equation in Models 10, 11 and 12 to predict the number of undos conditional on experiencing at least one undo. Results of sensitivity tests using generalized linear models with negative binomial link and grouped logits yield equivalent results. Editors $i = 30,272$; periods $t = 12$; total number of observations $= 212,317$. (Not all editors started editing in time period 1.) *p < .05 **p < .01 ***p < .001 (two-tailed tests)

To test **Hypothesis 1b**, we model the likelihood that editor i suffers a norm violation during time period t, DepVarit = Number of Times Editor i was Undoneit. Editor i may suffer no undos because he does not contribute to Wikipedia, or alternatively because no one undoes his or her edits even when they are numerous. In order to focus on the latter scenario, we again control for the possibility of the former in the three ways described above. Coefficient estimates on the density variable are in the same direction across all three methods, and for brevity we report only results using the Heckman-like correction. **Table 2.3** reports the results of these estimations. Consistent with Hypothesis 1b, we find in Models 7-9 that editors embedded in dense social networks are less likely to suffer an undo.

To test the robustness of our results, we checked whether the model can predict the number of times editor *i* was undone during time *t*, contingent on editor *i* being undone at least once during that period. To do so, we estimated *Was Undone at Least Once$_{it}$* = $y_{it}\mu + \tau_{it}$ where y_{it} includes a set of selection-independent variables, and then estimated *InverseMills$_{it}$* = $\phi(y_{it}\hat\mu)$ / $\Phi(y_{it}\hat\mu)$ where ϕ is probability density function, Φ is the cumulative normal density and $\hat\mu$ is the estimate of μ. We then include that *InverseMills$_{it}$* estimate in the random-effects negative binomial regression of *Number of Times Editor i Was Undone$_{it}$*. We report these results in Models 10-12 and obtain results directionally similar results to those in Models 7-9.[64]

---

[64] We also find that editors of articles edited by no one else were less likely to suffer an undo, and that those who co-edited with one other editor were no more likely to engage in undos than those who co-edited with two editors (once the number of articles was controlled for; see Models 9 and 12). Beyond that, an increase in the number of co-editors led to an increase in the likelihood of experiencing an undo. Likewise, the total number of articles edited by editor *i*, as well as his or her focus on a small number of articles, led to a higher incidence experiencing an undo. Editors who had signed up a long time earlier were more likely to experience undos, but that effect was offset by the negative effect of actually contributing edits to the project. Finally, editors who undid the edits of others were more likely to experience undos themselves.

Steps 2 and 3: Eliciting Norm Compliance and Compensating Those Who Elicit Compliance

To test **Hypotheses 2a** and **3a**, we modeled the likelihood that editor i reverts an undo during time period t. We distinguish between two types of reverts of undo: (1) editor i steps in to revert the undo of an article version saved by another editor, as captured by $DepVar_{it}$ = *Number of Times Editor i Reverted Others$_{it}$*, and (2) editor i steps in to revert an undo of an article version saved by another editor j who has at least once reverted an undo of another editor's work during (t-1), as captured by $DepVar_{it}$ = *Number of Times Editor i Reverted Others Who Reverted$_{it}$*.

       Table 2.4 reports the results of our estimations. In 13, 14, and 15 we examine the conditions under which, when another editor's article version is undone, editor i steps in to revert the undo. Across the three models, we find that editors embedded in high-density networks are not more likely to revert undos of another editor's work. This is inconsistent with **Hypothesis 2a**. In Models 16, 17 and 18, we examine the conditions under which editor i steps in to revert an undo of an article version saved by another editor j who had previously reverted an undo of another editor's work during (t-1), as captured by *Number of Times Editor i Reverted Others Who Reverted$_{it}$*. Across the three models, we find that editors embedded in high-density networks are more likely to engage in such behaviors and to reward those who had reverted others' work by reverting undos that affected them. This pattern of results supports **Hypothesis 3a**. Overall, this pattern of results suggests that Wikipedia editors in dense social networks do not blindly revert undos suffered by other editors. They only do that as a reward to others who engage in reverting undos for others.

       To test **Hypotheses 2b** and **3b**, we modeled the likelihood that editor i variables: (i) $DepVar_{it}$ = *Number of Times Editor i Was Undone Followed by No Revert$_{it}$*, (ii) $DepVar_{it}$ = *Number of Times Editor i Was Undone Followed by Editor i Reverts Undo$_{it}$*, and (iii) $DepVar_{it}$ =

Number of Times Editor i Was Undone Followed by Another Editor Reverts Undo$_{it}$.[65] Editor *i*

may of course experience no reverts simply because he did not suffer any undos, or because he

suffered undos but no one reverted them. As before, we control for the former possibility in three

ways. First, we exclude time periods *t* when editor *i* does not suffer any undos. Second, we

include a dummy variable equal to one when editor *i* suffered any undos. Finally, we use

Heckman-like correction, estimating the likelihood that editor *i* will suffer at least one undo

during time *t*. The specification for this function was given in model 9. Coefficient estimates on

the density variable are in the same direction across all three methods, and for brevity we only

report results using the Heckman-like correction.

Table 2.5 reports the results of our estimations. In Models 19 and 20 we examine the

conditions under which an article version saved by editor *i* was undone and followed by no

revert. Consistent with **Hypothesis 2b**, we find that editors embedded in high-density networks

are less likely to suffer an undo that is not followed by a revert. Consistent with **Hypothesis 3b**,

we find that editors who have reverted undos of other editors' work are less likely to suffer an

undo that is not followed by a revert and this effect is particularly strong if editor *i* is embedded

in high-density network.

In Models 21 and 22, we examine the conditions under which an article version saved by

editor *i* was undone and editor *i* personally reverted the undo. Consistent with **Hypothesis 2b**,

across both models we find that editors embedded in high-density networks are less likely to

revert an undo of their own work. Consistent with **Hypothesis 3b**, across the two models we find

that editors who have reverted undos for other editors are less likely to revert an undo of their

own work and this effect is particularly strong if editor *i* is embedded in a high-density network.

---

[65] *Number of Times Editor i Was Undone Followed by Editor i Reverts Undo$_{it}$* excludes situations in which editor *i* undid his own version and then reverted the undo.

Table 2.4. Negative Binomial Random-Effects Estimates that Editor *i* Reverted an Undo at Time *t* (Test of **Hypotheses 2a and 3a**)

Independent Variables	Number of Times Editor i Reverted Others$_{it}$			Number of Times Editor i Reverted Others Who Reverted$_{it}$		
	Model 13	Model 14	Model 15	Model 16	Model 17	Model 18
Network Density$_{it-1}$	-.09	-.11	-.05	.15*	.16*	.17*
	(.09)	(.09)	(.09)	(.07)	(.07)	(.07)
Network Size$_{it-1}$	.09***	.02	.09***	.07***	.08***	.08***
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)
Network Size0$_{it-1}$	.28***	.47***	.34***	-.09	-.19	-.04
	(.09)	(.18)	(.12)	(.20)	(.17)	(.20)
Network Size1$_{it-1}$	.32*	.07	.28*	.27	.26*	.33*
	(.13)	(.13)	(.13)	(.15)	(.12)	(.15)
Percentage of Articles Edited More than Twice$_{it-1}$	-.17	-.08	-.34***	-.11	-.06	-.11
	(.10)	(.11)	(.10)	(.11)	(.11)	(.11)
Cumulative Edits$_{it-1}$	—	.11***	.19***	—	.02	.05
		(.02)	(.02)		(.11)	(.11)
Months since Signup$_{it-1}$	—	-.56***	-.37***	—	.08	-.03
		(.12)	(.11)		(.09)	(.09)
Number of Edits$_{it}$x10	.01***	.01***	-.01***	.47*	.47**	.37
	(.003)	(.003)	(.003)	(.19)	(.16)	(.19)
Time Period Dummies$_t$	No	No	Yes	No	No	Yes
Inverse Mills Ratio$_{it}$	No	No	Yes	No	No	Yes
-Log-Likelihood	25,731	25,401	24,943	20,394	20,073	19,994
Degrees of Freedom	6	8	20	6	8	20
Wald χ^2	12,365	12,466	12,598	12,336	12,367	12,659

Note: Numbers in parentheses are standard errors. Constant was omitted. All χ^2 tests are based on a baseline model with no covariates. Selection equation predicting an undo of a version last saved by editor *i* based on previous experience, tenure and month dummies was omitted. Resulting Inverse Mills ratio was used as control in the outcome equation, as noted. Results of sensitivity tests using generalized linear models with negative binomial link and grouped logits yield equivalent results. Editors *i* = 30,272; periods *t* = 12; total number of observations = 212,317 (not all editors started editing in time period 1). As before, results are stable with respect to risk set. *p < .05 **p < .01 ***p < .001 (two-tailed tests).

Table 2.5. Negative Binomial Random-Effects Estimates that Editor *i* Experienced a Revert of an Undo at Time *t* (Test of **Hypotheses 2b and 3b**)

Independent Variables	*Number of Times Editor i Undone Followed by No Revert$_{it}$*		*Number of Times Editor i Was Undone Followed by Editor i Reverts Undo$_{it}$*		*Number of Times Editor i Was Undone Followed by Another Editor Reverts Undo$_{it}$*	
	Model 19	Model 20	Model 21	Model 22	Model 23	Model 24
Network Density$_{it-1}$	.28***	-.17***	-.34**	-.93***	.31**	.19**
	(.04)	(.04)	(.14)	(.14)	(.08)	(.08)
Number of Times Editor i Reverted Others$_{it-1}$	-.12***	-.12***	-.22**	-.27**	.18**	.14**
	(.02)	(.02)	(.05)	(.05)	(.04)	(.04)
Number of Times Editor i Reverted Others$_{it-1}$ * Network Density$_{it-1}$	-.02**	-.02**	-.06**	-.05**	.11**	.12**
	(.01)	(.01)	(.01)	(.01)	(.03)	(.03)
Network Size$_{it-1}$	.44***	.35***	.93***	.47***	.72***	.59***
	(.01)	(.01)	(.03)	(.03)	(.02)	(.02)
Network Size0$_{it-1}$	.26***	-.50***	-1.49*	-1.80***	.55***	.11
	(.05)	(.05)	(.22)	(.22)	(.10)	(.09)
Network Size1$_{it-1}$	.46***	-.02	-.62	-.95**	.75***	.30*
	(.07)	(.07)	(.32)	(.32)	(.15)	(.15)
Cumulative Edits$_{it-1}$	-1.35***	-.46***	-.90***	-.21**	-.66***	-.80***
	(.05)	(.02)	(.17)	(.07)	(.10)	(.04)
Months since Signup$_{it-1}$	1.16***	.28***	.77***	.40**	.62***	.70***
	(.05)	(.02)	(.16)	(.06)	(.09)	(.04)
Number of Times Editor i Was Undone Followed by No Revert$_{it}$	—	—	.04***	.03***	.04***	.04***
	—	—	(.00)	(.00)	(.00)	(.00)
Number of Times Editor i Was Undone Followed by Editor i Reverts Undo$_{it}$	.04***	.03***	—	—	.03***	.00
	(.00)	(.00)	—	—	(.00)	(.00)
Number of Times Editor i Was Undone Followed by Another Editor Reverts Undo$_{it}$	.07***	.06***	.11***	.05***	—	—
	(.00)	(.00)	(.01)	(.01)	—	—
Number of Times Editor i Undid Others$_{it-1}$	—	.31***	—	1.16***	—	.42***
	—	(.01)	—	(.02)	—	(.01)
Edits$_{it}$ x10	.27*	.15**	.23	.19	.37	.18
	(.10)	(.10)	(.33)	(.56)	(.32)	(.32)
Time Period Dummies$_t$	No	Yes	No	Yes	No	Yes
Inverse Mills Ratio$_{it}$	Yes	Yes	Yes	Yes	Yes	Yes
-Log-Likelihood	110,547	109,923	20,211	18,384	36,409	36,153
Degrees of Freedom	14	26	14	26	14	26
Wald χ^2	32,173	32,880	6,454	11,157	10,332	11,802

Note: Numbers in parentheses are standard errors. Constant, number of articles, and percentage of articles edited more than twice were omitted. All χ^2 tests are based on a baseline model with no covariates. Selection equation

Finally, in Models 23 and 24, we examine the conditions under which an article version saved by editor i was undone and then reverted by another editor. Consistent with **Hypothesis 2b**, across the two models we find that editors embedded in high-density networks are much more likely to experience an undo followed by a revert by another editor.[66] Consistent with **Hypothesis 3b**, we find that editors who have reverted undos for other editors are more likely to experience an undo followed by a third-party revert and this effect is particularly strong when editor i is embedded in a high-density network.

Continued Participation

To test **Hypotheses 4a**, **4b** and **4c**, which pertain to continued editor participation, we modeled the likelihood that editor i contributes at least one edit during time t. The dependent variable was coded one if the editor has contributed at least once during time t and zero otherwise. To estimate this model, we used a fixed-effects panel logistic, with joint probability function given by:

$$\rho_{it} = \frac{1}{1 + e^{-\beta x_{it}}} \qquad (2.4)$$

Table 2.6 presents the results of Models 25, 26 and 27. Consistent with our expectations, we find across the three models that editors surrounded by dense network structures are more likely to continue contributing content to Wikipedia. In Model 25 we find that having one's contribution undone has on average no effect on the likelihood of continuing to write for

---

[66] We have also run auxiliary models in which we took the dependent variable from Models 19 and 20 and split it in two. First, we examine the conditions under which an article version saved by editor i was undone and then reverted by an editor with whom editor i has previously worked. Second, we examine the conditions under which an article version saved by editor i was undone and then reverted by another editor with whom editor i has not previously worked. Consistent with our expectations, the effect of density on the likelihood of a revert by another editor is higher if that editor has previously worked with editor i.

Wikipedia. Model 26 however, reveals a great deal of variation in the effect of an undo on the likelihood of continued participation, depending on whether the undo was reverted and, if so, how. In contrast to the predictions of **Hypothesis 4a**, the results indicate that an undo left unreverted has no effect on the likelihood of continuing to contribute to Wikipedia.[67] Consistent with **Hypothesis 4b**, however, the results indicate that an undo personally reverted by the editor whose work was undone makes him or her less likely to continue contributing. Furthermore, consistent with **Hypothesis 4c,** the results indicate that an undo reverted by a third party makes the editor whose work was undone more likely to continue to contribute.

For completeness, in Model 27 we also interact with the three variables associated with the three hypotheses with a measure of density around actor i at time t. We find that when editor i is surrounded by a dense network, the effect of unreverted undos remains the same. However, when editor i is surrounded by a dense network, the effect of undos of editor i's work subsequently reverted by editor i is even more negative. This should not be surprising. An editor surrounded by a dense network is likely to expect that the network will revert the undo on his or her behalf. Failure of the network to do so, requiring the editor to step in and personally revert the undo, makes him or her more disappointed with the network and thus more likely to leave.

In contrast, when editor i is surrounded by a dense network, the effect of undos of editor i's work reverted by another editor is even greater. This too should not be surprising. An editor surrounded by a dense network is apt to expect the network to revert the undo on his or her behalf. The expectation that the network surrounding him will continue to do that in the future makes the editor less likely to leave.

---

[67] We suspect that the lack of statistical significance occurs because some reverts go unnoticed by the editor, and thus are unlikely to have an effect on the editor's editing pattern.

Table 2.6. Fixed-Effects Logistic Estimates that Editor i Makes at least One Edit during Time t (Test of **Hypotheses 4a-c**)

Independent Variable	Model 25	Model 26	Model 27
Network Density$_{it-1}$	.08*	.08*	.07*
	(.04)	(.04)	(.03)
Number of Times Editor i Was Undone$_{it-1}$	.00		
	(.01)		
Number of Times Editor i Was Undone Followed by No Revert$_{it-1}$		.02	.01
		(.01)	(.01)
Number of Times Editor i Was Undone Followed by Editor i Reverts Undo$_{it-1}$		-.10***	-.07**
		(.02)	(.02)
Number of Times Editor i Was Undone Followed by Another Editor Reverts Undo$_{it-1}$		.08**	.04**
		(.03)	(.01)
Number of Times Editor i Was Undone Followed by Undo Followed by no Revert$_{it-1}$*Density$_{it-1}$			.08
			(.05)
Number of Times Editor i Was Undone Followed by Editor i Enforces Reverts Undo$_{it-1}$*Density$_{it-1}$			-.66**
			(.23)
Number of Times Editor i Was Undone Followed by Another Editor Reverts Undo$_{it-1}$ * Density$_{it-1}$			.33*
			(.16)
Number of Times Editor i Undid Others$_{it-1}$	Yes	Yes	Yes
Percentage of Articles Editor i Edited More than Twice$_{it-1}$	Yes	Yes	Yes
Network Size$_{it-1}$, Network Size0$_{it-1}$ and Network Size1$_{it-1}$	Yes	Yes	Yes
Number of Articles Edited$_{it-1}$	Yes	Yes	Yes
Edits$_{it-1}$*10	Yes	Yes	Yes
Cumulative Edits$_{it-1}$	Yes	Yes	Yes
Months since Signup$_{it-1}$	Yes	Yes	Yes
Periods since Last Edit$_{it-1}$	Yes	Yes	Yes
Time Period Dummies$_t$	Yes	Yes	Yes
-Log-Likelihood	52,421	52,311	51,403
Degrees of Freedom	23	25	28
Wald χ^2	24,188	24,209	24,226

Note: Numbers in parentheses are standard errors. Constant was omitted from table. All χ^2 tests are based on a baseline model with no covariates. Editors i = 19,290; number of periods = 12; total number of observations = 153,582. (Not all editors started editing in time period 1.) The number of editors is smaller than in previous tables because fixed-effects estimation removes all editors who always edited or never edited from the risk set. *p < .05 **p < .01 ***p < .001 (two-tailed tests)

The results we present provide overall support for our hypotheses, but they have some shortcomings. First, we do not directly measure relationships between individuals, such as social exchanges of messages between editors. Instead, we infer the existence of relationships by identifying who worked with whom on a given article. We believe, however, that this shortcoming does not undermine our results, and indeed that it makes our results conservative. Consider what would happen if we erroneously assumed that relationships exist when they do not—in other words, that a network of editors is dense when in reality it is not. We would expect these editors to behave in the manner described by the theory, but because there is no density between them they would not do so. As a consequence, we would be less likely to obtain the results we do. Conversely, we may have mistakenly assumed that relationships do not exist when in reality they do. In this scenario, we would be underestimating the extent of density between editors—in other words, we would not expect these editors to behave as described by the theory, though in fact they do. This scenario too would make it less likely that we will find the results we do. Both of these measurement errors suggest that our results are fairly conservative estimates.

We also labor under the disadvantage of being unable to measure all types of norm violations. This would be a problem if, for example, editors in dense networks were less likely to undo edits, but more likely to violate norms on, say, the talk pages where editors discuss how an article should evolve. If this were the case, however, we would expect extensive spillovers, such that editors who violate norms on, say, talk pages would be more likely to experience retribution in the form of undos of their article versions. This scenario should lead to a positive relationship between density and the likelihood of experiencing undos, making it less likely that we will

observe a negative relationship between the two. Thus the negative relationship we document should be seen as a conservative estimate.[68]

We are also unable to capture all types of punishment of norm violations. For example, some editors who undid articles might have been punished via private e-mails. This scenario could present a problem for interpretation of our results in the following way. Suppose no real relationship exists between density and reverts of undos, but editors do tend to chastise other editors embedded in sparse networks via private communication, and those embedded in dense networks via public reverts of their undos. If this differential treatment existed, we would observe a relationship between density and reverts of undos even if it did not exist. It is very unlikely, however, that this scenario actually prevails. If anything, we would expect editors embedded in dense networks to be less likely than those with sparse networks to have their undos reverted (for fear of retaliation, say). Thus this potential bias makes the results we observe less rather than more likely. Finally, it is unlikely that we capture all types of rewards for those who punish norm violations. Such rewards can take the form of private thank-you e-mails, public thank-you entries on editors' private pages, and other expressions of gratitude. Once again, to the extent that such rewards are substitutes for reverts of undos, we should be less likely to observe the results we do.

Finally, there remain the issues of reverse causality and unobserved heterogeneity. With respect to reverse causality, it is possible for an individual editor to create a high-density network around himself or herself by introducing acquaintances to each other. In gratitude for such introductions, the acquaintances may in turn refrain from violating norms against the editor, or

---

[68] Similar logic could be applied to an unobserved propensity of editors in high-density networks to experience norm violations on, say, talk pages. Again, to the extent that editors penalize such behavior via undos, we should observe a positive rather than negative association between density and undos.

may punish those who do so. Though this scenario could occur, we have undertaken a number of steps to exclude it from the data. Suppose editor i works with editor j on article A, and with editor k on article B. Editor i may tell editor k about his work with editor j and invite him to join the two of them in working on article A, which editor k does. Because our definition of network density around editor i explicitly requires editors j and k to work together on a different article in which i does not participate, those two editors would then have to start editing another article, C. They would also have to attribute this new undertaking to editor i's introduction, and in gratitude perform fewer undos or revert more undos affecting editor i. We doubt that such joint editing activity on article C would be attributed to editor i, and thus we do not believe that reverse causality is responsible for our results.

Concerns about unobserved heterogeneity can also lead one to argue that editors engaging in or suffering from fewer norm violations; engaging in or witnessing more punishments of norm violators; and rewarding those who punish norm violations, as well as getting rewarded for such acts, find themselves in this situation not because of density, but because of their unobserved personal characteristics. A critic can then argue that these unobserved characteristics are correlated with editor's proclivity to form dense networks, which results in the empirical association between density and the six types of behaviors described above. We believe that these concerns are attenuated by the fact that auxiliary analyses we ran using fixed effects models (see footnote 63 and examine estimation procedure for models 25-27) generate similar pattern of results, imply that the time invariant unobserved characteristics cannot be held responsible for generating the results. The unobserved heterogeneity explanation of our results is thus limited to only the time-varying unobserved effects.

CONCLUSIONS

Since the inception of the discipline, sociologists have examined the role of dense social relationships in various social phenomena. People surrounded by friends who are also each other's friends are thought to enjoy more social and economic support (Durkheim 1951; Uehara 1990). They are also believed to be less likely to commit or suffer norm violations. Coleman (1990) formalized this intuition and argued that high-density networks enable third parties to compensate norm enforcers for the expense of chastising norm violators. Such payments encourage actors to punish those who violate norms, in turn reducing the incidence of norm violation. Despite ubiquitous citations of Coleman's explanation, little empirical work has tested it convincingly. This is problematic; we do not know whether the mechanism is borne out in reality. If not, we may erroneously recommend that a network be made denser even if doing so will not improve norm enforcement. Our paper endeavors to address this issue by testing Coleman's mechanism in detail. We find substantial support for it, suggesting that increasing network density to elicit norm compliance is justified. Support for Coleman's mechanism alerts us to the importance of punishments for norm violations and rewards for such punishments, and thus helps us design social systems in which norms are observed.

The fact that we found supporting evidence in the Wikipedia context highlights a number of conditions that promote the operation of Coleman's mechanism. Ln Wikipedia, for example, norm violations, punishments of norm violations and as rewards for punishing norm violators are all highly visible. Replicating these conditions in the design of a social system is critical; otherwise norm violations will remain undetected and therefore unpunished. Wikipedia's norms are also clearly articulated, making it easy to detect a violation and fairly difficult to claim that a norm violation occurred when it did not. It is also reasonably clear how to punish violators in ways that will elicit rewards from others. Without such clear specification of appropriate

punishment, some actors may be afraid to administer it for fear of committing a violation themselves.

Understanding such conditions has important implications for related streams of the literature, such as the effort to link network density to performance. On the one hand, higher network density is believed to constrain the novelty and creativity of new ideas and solutions, and thus individual and collective performance (Burt 2005). On the other hand, higher network density is thought to enhance performance via a higher rate of norm compliance (Uzzi 1999). Numerous papers seek to address this tradeoff by pointing to sets of conditions under which one or the other effect is likely to be stronger, suggesting for example that performance will be higher in dense networks when tasks are collective and require everyone's cooperation (Ahuja 2000). Our results indicate that this positive association will hold only when mechanisms for punishment and reward of punishment are in place. Otherwise, dense networks will suffer all the shortcomings of constrained creativity without enjoying any of the benefits of higher norm compliance.

The theory and results we present here also inform our understanding of what makes social systems survive. Specifically, they underscore a tradeoff in designing a social system between maximum norm compliance and maximum longevity. Our results indicate that norm violations followed by punishments make people more committed to a social system than they would be if they had never experienced a norm violation. To the extent that very dense networks discourage norm violations, they also prevent actors from learning just how strong the community is. Even a small decrease in network density will increase the rate of norm violation, whose punishment will in turn promote greater commitment to the social system. These conclusions are similar in nature to those of Uzzi (1999), who found that intermediate levels of

density promote the highest performance. Whereas Uzzi's quantitative findings pertain to performance, our paper quantitatively estimates commitment to a social system.

We hope that our paper will stimulate further research. We see substantial opportunities for further tests of Coleman's mechanism. Specifically, it would be helpful to document the conditions under which network density has no effect on norm enforcement. If Coleman's theory is correct, for example, it should be the case that when norm violations, punishments and rewards for norm punishments are hard to observe, density will have limited effect on these phenomena. Density should also have no effect on populations of individuals who derive sufficient intrinsic rewards for punishing those who violate norms. Furthermore, density will not lead to norm observance when rewards for punishment are very expensive to provide, such that a third-order free riding problem occurs. Finally, further opportunities exist to show that density may actually reduce the incidence of norm observance. This mechanism would be most likely if punishing friends who are also each other's friends were particularly costly. Demonstrating that the relationship between density and norm observance critically depends on such factors would further lend credence to Coleman's theory.

We hope that future research will take advantage of the vast amounts of data on social interactions on the internet. This unprecedented opportunity for insight into human interactions makes it possible to offer unequivocal empirical support for many theories central to sociology. For example, a string of papers using e-mail data has convincingly shown that homophily, as distinct from other mechanisms, does indeed explain why actors with similar characteristics are more likely to form relationships with each other (Kossinets and Watts 2009; Menchik and Tian 2008). This paper too provides empirical support for a widely accepted mechanism. It is to be hoped that future papers will furnish unambiguous evidence for other widely cited social

theories. We hope too that a new set of papers will take advantage of the fact that on-line environments make certain social mechanisms more salient, allowing for development of new theories. For example, Piskorski (2010) has shown that on-line social networks allow people to create an illusion of constant sociability, which they can then use to engage in other, often illegitimate, activities. Similarly, one can argue that such social platforms make others' patterns of social relationships public information, in turn illuminating opportunities for individuals to act as social brokers. Viewed as such, on-line environments can help us further our theories of brokerage. Other theory-development opportunities abound.

In the next chapter, I examine the processes through which individuals become involved in, contribute, and leave Wikipedia, with particular attention to the relationship between turnover and the social structure of online participation in the context of individuals' overall social networks. This question is addressed using data from in-depth interviews with thirty-five current and former Wikipedia contributors, coupled with a brief quantitative analysis of reported offline social networks and socio-demographic characteristics and publicly available Wikipedia contribution characteristics. While the present study suggests that dense online social networks are beneficial because they reduce the likelihood that individuals are affected by norm infringement and increase the likelihood that their contributions are defended, the findings from the last chapter suggest that forging social connections with online collaborators to the detriment of real-life involvement can be detrimental to long-term participation in Wikipedia collective production.

WHEN COLLEAGUES COUNT, BUT NOT TOO MUCH:

SOCIAL NETWORKS AND TURNOVER MECHANISMS IN WIKIPEDIA

INTRODUCTION

Why do some individuals continue contributing to joint work on projects while others slack or

give up despite dealing with the same environment? Why is it that, given the same organizational

environment, some individuals seek to manage the expectations and pressures of collaborating,

and remain committed to the organization while others fail to do so and exit the organization?

Organizational scholars have considered the question of motivation for participating in various

collaborative settings, and of satisfaction with one's participation, but overall our ability to

explain variation in levels of organizational commitment remains rather limited (Randall 1988).

Many studies measuring organizational commitment as result of individual motivation rely on

one-time individual level survey data and interviews with participants, in which intrinsic and

extrinsic motivations are reported as reasons for engaging in collaborative work (Deci and Ryan

2000). Few studies recognize that motivations may, and often do, change during participation on

a project and therefore they are not a reliable predictor of individual's satisfaction or turnover

patterns (Freeman 2007).

Two main sources of work satisfaction and antecedents of commitment in collaborative

work are one's task-related activities and one's interactions with other organizational members

(Leiter and Maslach 1988). Researchers have shown that interaction patterns and interpersonal

conflicts are predictors of organizational commitment but did not extensively specify the

underlying dynamics of these processes (Morris and Sherman 1981; Eisenberg, Monge et al.

1983). Similarly, little attention has been paid to the processes through which the characteristics

of individuals' overall social networks affect their capacity to commit to an organization they participate in.

This study juxtaposes research on social networks and power-dependence theory to examine individual commitment as reflected by turnover decisions, and proposes that high embeddedness[69] in social networks of collaboration is detrimental to individual commitment under conditions of high dependence on these networks. Predicated on the belief that individual turnover decision can be understood through one's narrative of participation, I use semi-structured interviews based on Atkinson's "life story" interview method (1998) to elicit information about contributors'[70] trajectories of participation in the writing of English Wikipedia articles, with particular attention to interpersonal dynamics and social networks of article contributors. I then analyze the retrospective accounts of participation and triangulate them with objective participation information retrieved from archival data in order to ascertain the existence of a relationship between the network of co-participation and the place of this network in the individual's overall social network, and the turnover decision.

The type of setting chosen for this study, a novel form of online volunteer work which emerged as a result of technological support for collective action behavior (Benkler 2006) is crucial in identifying the link between social networks and participant turnover. Because collective production of Wikipedia articles is asynchronous and involves remote collaboration with other contributors, interviewees' narratives of volunteer "careers" are unencumbered by the

---

[69] In this study, embeddedness is used to define an individual position in the structure of production where the individual has multiple ties to the same group of which he or she is a member. Thus, an individual within a dense network of coworkers who has multiplex ties (such as work, advice, friendship) to several coworkers would be considered embedded, while someone in a dense work network whose social ties are situated outside the network would not be classified as embedded.

[70] Throughout this study I use "participant", "editor" and "contributor" interchangeably to refer to individuals involved in the process of writing Wikipedia articles.

complex embedded "web of relationships" (Simmel 1964) that characterizes everyday life and by

the complex rewards associated with workplace settings. This aspect allows me to explore how

individuals' own choices of allocating time and effort to online article contributions and to

relationships with other contributors affect their commitment to Wikipedia and their likelihood

of turnover. In the following section, I briefly review research on social networks, conflict and

turnover in collaborative production settings as a source of research questions and comparisons

prior to data analysis (Strauss 1987; Strauss and Corbin 1998).

Social Networks and Turnover

Social networks represent relationships arising among a given set of individuals as a result of

personal interactions and co-memberships. Social networks facilitate advice (Sparrowe, Liden,

Wayne and Kraimer 2001) and sharing of information (Burt 2000) but also serve as identity

referents for individuals to learn about participation in an organization (Bolino, Turnley, and

Bloodgood 2002). For example, individuals can learn through their social networks the social

norms surrounding contribution and collaboration, as well as the performance criteria and

expectations associated with their organizational role (Mossholder, Settoon, and Henagan 2005).

Individual outcomes such as promotions (Burt 1992) and turnover (Krackhardt and Porter

1985) are influenced by positions within social networks. More recently, researchers such as

Kahn (1998) and Burt (2001) have identified mechanisms through which employee ties may

affect their attachment to an organization. Kahn (1998) has argued that lack of ties with other

employees may lead to turnover due to lack of emotional engagement. From a structural

perspective, Burt (2001) has found that more embedded participants have higher organizational

attachment than individuals who are less embedded in an organization, a finding consistent with

the fact that individuals with more ties within an organization are those least likely to exit it (McPherson, Popielarz, and Drobnic 1992).

While research has examined the relationship between individual networks within the organization and turnover, less concern has been granted to the size and structure of one's organizational social network relative to one's overall social network with the exception of work-family (or, work-life) research. In the next section I summarize research work-family balance as it pertains to the topic of social networks and turnover, before presenting my research setting and data collection effort.

Social Networks within Organizations in the Context of One's Overall Social Network

Organizational membership represents a fundamental aspect of everyday life, and hence a large majority of organizational research on social networks has concerned itself with social networks within organizational boundaries. At the same time, the organizations we work for are only one of many social settings in which we interact with each other and form relationships. However given the preponderance of dual-career families and increase in average times spent at work (Ilies et al. 2007), organizational researchers have become increasingly concerned with questions regarding the balance between social networks and roles at the workplace and in personal life.

"Attention to the balancing of work and family roles has traditionally focused on conflict or interference between these roles (Eby et al. 2005). Work-family conflict occurs when pressures from the work and family domains are mutually incompatible" such as when time, strain, or specific behaviors associated with one domain hinder individual performance in the other (Illies et al. 2007: 1368). While research has not directly linked role conflict with turnover, meta-analyses of work-family conflict indicate that individuals in this situation are likely to have

lower organizational attachment and commitment (Mesmer-Magnus and Viswesvaran 2005) than other employees which suggests a higher likelihood of organizational turnover.

At the same time, "positive spillovers [between work and family domains] have traditionally been neglected" in the literature (Ilies et al. 2007). While the psychological literature offers a statistical examination of positive affect spillovers between work and family (Ilies et al. 2007), this study proposes a qualitative analysis of a volunteer collective production setting to identify a social mechanism through which personal networks can play a positive role in reducing organizational turnover.

RESEARCH SITE

This study relies on data from the online encyclopedia Wikipedia, a free online encyclopedia collaboratively created by people around the world. Because anyone may contribute to Wikipedia, and contributions are not screened or censored before becoming part of the encyclopedia,[71] Wikipedia has attracted to date over six million registered contributors who created over 3.5 million articles in English, and a total of over 16 million articles in over 250 languages. Due to its broad coverage of knowledge Wikipedia is ranked as the 6th most visited website in the world, according to Alexa Traffic Rank 2012.[72]

The success of collective article writing in Wikipedia relies on a technology named 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

---

[71] Before late 2008, according to information retrieved on December 21, 2010 from http://en.wikipedia.org/wiki/Wikipedia:Flagged_protection_and_patrolled_revisions, in *Wikipedia: The Free Encyclopedia*, Wikimedia Foundation Inc. [Encyclopedia on-line].

[72] Data retrieved on January 25, 2012 from http://www.alexa.com/topsites.

accessible through a history page.[73] Participants disagreeing with page changes may alter or

erase these changes in response. Wiki software represents a radical change in collective

production philosophy compared against the classical idea of bulletin boards where people may

contribute only by accretion and existing text cannot be eliminated by peer participants.[74]On

Wikipedia contributors thus forsake ownership and control over their work, trading off

individual visibility for collective work, and risking that their contribution is modified or

eliminated by peers. In this sense, Wikipedia article writing is similar to commonly used forms

of offline collective production of a presentation or a scientific research project, except that the

main contributors are ordinarily listed as such in the latter.

Wikipedia is structured as a website, with new pages being continually added and linked

into the existing structure. Every Wikipedia article consists of a set of three interrelated pages:

Article page, *Discussion* page, and *History* page. The Article page displays the contents of the

most recent version of that respective entry. To make changes to this entry, contributors may

click on an "Edit Page" button, which presents them with an editable version of the contents of

the article page. This allows any reader interested in contributing to modify the content in

various ways - such as making large contributions, copy-editing text, or adding references and

photographs to improve an article. Wikipedia's software platform provides a complete history of

any given article. Anyone may find summary information[75] about a contributor's edits or about

---

[73] The description of wiki software here closely matches its implementation by the Wikimedia Foundation, the non-profit legal entity behind Wikipedia. Various Wikimedia software implementations vary in their feature set.

[74] It is technically possible to erase comments but the deletion action is very rare on Wikipedia, and it takes place only when exposure of sensitive personal information such as phone numbers, or social security numbers is at stake. The action of locking article pages is slightly more frequent, but it took place for less than 0.05% of articles in my dataset.

[75] This summary information contains the time and date of edit, user name/ IP address, as well as a link to the former's talk page, to a list of the registered user's contributions, the size of the edit, and a brief note regarding the nature of the edit provided by the editor or generated automatically by a Wikipedia script.

the history of any given article by accessing the History page, and examine the difference between two versions of the same page. In addition to the article pages, Wikipedia contains several types of infrastructure pages such as community rules, manuals of editing and writing style, pages listing articles organized within various *WikiProjects* (e.g., History of France, Physics), and statistics about Wikipedia. The Wikipedia page space is tightly linked to the *Help*, *Special* (pages), and *Template* spaces, which contain tools facilitating socialization within the project, administration and format standardization of project pages. *User* (contributor) pages are personal spaces for registered editors' self-expression, containing free-format information about the interests, opinions, identity, or activities that the owner wants to share with the community.

Discussion pages are never independent of other pages but always attached to other pages such as articles or user pages. Discussion pages offer space for discussing and debating the content of the primary page, asking for information, or leaving messages for other contributors. Discussion pages are also used for talking about the evolution of Wikipedia and for proposing various changes in the structure, style or norms applied to the editing process. Formal Wikipedia contribution and collaboration norms have gradually crystallized as a result of such public discussions.

The first formally recorded Wikipedia rule, *Neutral Point of View* emerged during 2001, and was later acknowledged as the foundational principle of the Wikipedia community.[76] More rules and guidelines followed soon after, and now Wikipedia policies are targeted towards two main facets of article production: writing criteria and collaboration rules. The former clarify the

---

[76] "Wikipedia Timeline," in *Wikipedia: The Free Encyclopedia*, Wikimedia Foundation Inc. [Encyclopedia on-line]. Available from http://meta.wikimedia.org/wiki/Wikipedia_timeline; retrieved January 2, 2008, and Wales, James. "Statement of Principles" in *Wikipedia: The Free Encyclopedia*, Wikimedia Foundation Inc. [Encyclopedia on-line]. Retrieved on December 21, 2010 from http://nostalgia.wikipedia.org/w/index.php?title=Jimbo_Wales/Statement_of_principles&oldid=75340

purpose of Wikipedia as a repository of unbiased, accurate and verifiable synthesis of expert knowledge, while the latter set the code of conduct and social rules governing the behavior of contributors to the project. On Wikipedia, contributing in "good faith" and assuming good faith on behalf of others, as well as fostering an environment of open and enthusiastic contribution and benevolence are upheld against making 'perfect contributions'. One is required to recognize the equality of contribution rights and to expect contributions to be altered in the interest of article improvement.

In addition to collaboration norms, Wikipedia participants have set up guidelines and institutions such as mediating committees to ensure contributions are evaluated fairly and without personal bias. The concepts of community and collaboration norms are at the core of my interview data collection. In the next section I explain my data collection efforts, including my participation to Wikipedia and the steps I have undertaken in designing an interview protocol, selecting participants, and conducting the interviews.

DATA COLLECTION

My interaction with Wikipedia extended over a period of five years between 2006 and 2011 during which I observed contributor behavior and interactions, participated in article writing, and had formal and informal conversations with participants to the English Wikipedia. In order to understand the relationship between social networks and participation in this setting I have relied primarily on data from thirty-five semi-structured interviews in which English Wikipedia former or current participants who have narrated the history of their contributions to the free online encyclopedia. Although the interviews themselves offered a detailed picture of the social processes at play in participation to Wikipedia, I have also spent time as participant observer in the Wikipedia environment in order to gain a fuller understanding of the participation narratives.

114

Additionally I have engaged in archival research to retrieve objective data on interviewees'

Wikipedia contributions in order to mitigate potential interviewee recall biases regarding

information such as the date when they started contributing, or the frequency of their

contributions.

The semi-structured interview schedule included in **Appendix B** touches upon four broad

areas of interviewees' participation story in such an order that disruptions to conversation flow

and interviewees' frame of mind are minimized (Weiss 1994). The interviewees were

encouraged to follow a loose chronological progression: they were asked about the

circumstances of their first contributions, and then about strategies they employ during article

writing, about interactions with other participants, positive and negative experiences, and sense

of community. In the end, participants were asked to report socio-demographic characteristics,

educational background, and occupation, as well as a description of their personal social

networks.

Thirty of these interviews were collected by the author, with the help of two additional

researchers, between July and December 2008; five more interviews were collected during the

spring of 2011. Eighteen interviews were conducted on the phone and lasted between 55 and 90

minutes; thirteen were conducted via email correspondence, and four via instant messenger,

online real-time text-based discussion. Interviewees were a rather diverse set of individuals,

ranging from high-school students to retired professionals, from individuals working towards

their GED to individuals with several graduate degrees, and from Americans to individuals living

in the United Kingdom, Sweden, Serbia, Singapore or Australia.[77]

---

[77] Interviewees had much less gender diversity: only one of the interviewees was female, and another one refused to specify a gender.

The interviewees were selected through theoretical sampling from a list of contributors with over one hundred article contributions, where the main sampling category was based on the outcome of interest: continued participation versus turnover. A total of one hundred and forty-nine participants were randomly contacted using lists available on the Wikipedia site. Information about their total number of contributions and total number of articles they contributed to, and whether they are still contributing or have exited was collected from archival sources. In order to account for the possibility that individuals who belong to different cultures or cohorts of contributors (based on their starting dates) might have different participation "careers" I collected data to account for these variations as well. I started analyzing the data as the interviews were unfolding, and I expanded the sample during the data analysis stage until it reached theoretical saturation. The final sample consisted in approximately equal numbers of interviewees who are active, who have decreased the intensity of their participation, and who have ceased participating in Wikipedia work.

DATA ANALYSIS

As soon as the process of interviewing participants was under way, I proceeded by coding transcripts and by iteratively revising the interviewee sampling, the interview schedule, and the coding scheme. In doing so I have followed Strauss' four basic guidelines: asking the data a specific and consistent set of questions regarding individual process of participation in Wikipedia, analyzing the data carefully in order to ground my theory, writing notes and memos pertaining to my findings, and avoiding assumptions regarding certain patterns of participation until they emerged from the data (1987). These four main activities of grounded analysis – data-collection, note-taking, coding and memo'ing – were conducted simultaneously. The open coding process was done using both in-vivo concepts that emerged from participants' own

stories and concepts based on analytic categories induced by the semi-structured interview topics. After open-coding the first half-dozen interviews, three core code categories have started to emerge. These categories refer to online participation (including interactions and relationships with other participants, and contributor activities and one's role on the project), satisfaction / dissatisfaction with Wikipedia and decision to stop contributing, and interviewee's offline social network and context (see **Table 3.1** for a hierarchy of codes). I encouraged interviewees to speak about the relevance that these emerging themes and coding categories have for their commitment to Wikipedia. While doing so, I also used the initial coding as a 'springboard' to speculate about patterns and explore the interview data instead of staying bound to my initial findings (Strauss 1987).

I subsequently started using the initial categories as coding frames to successively sort cases into various classes. Since the initial question centered on commitment to participating on Wikipedia, a first subdivision classified participants as *highly-active* (coded S for *stayers*), *occasional participants* (coded P for *peripheral*)*, and *departed* (coded L for *leavers*). The next subdivisions emerged as a result of axial or intensive, focused coding around the categories of interest (Strauss 1987). Axial coding is a procedure which allows for increasing specification of conditions, consequences, interactions and strategies associated with a phenomenon of interest, committed participation, and for relationships among codes to emerge (Strauss 1987). In this case, each of the three initial categories of participants was divided into two classes, *highly embedded* and *low social participation* to the project. An additional attribute allowed me to separate interviewees by type of offline environment into socially and professionally *active* people versus *under-involved* interviewees.

117

The last coding stage consisted of a selective coding process whereby subordinate codes were consolidated, eliminated or linked to the main categories in order to shape the theory. This process resulted into clusters of codes focused on *online tasks*, *perceptions of online relationships*, *existence and enforcement of norms*, and *offline relationships*, all of which were then related to the core code of *participation*.

Table 3.1. Code Hierarchy

Code families	# Codes	Code categories
Online embeddedness: relationships and roles	36	Existence and enforcement of norms (21) Rewards and status (2) Time spent (5) Perceptions of online relationships (8)
Offline relationships and context	26	Talking with friends about Wikipedia (3) Education (3) Job (6) Friendships (8) Family status (4) Location (2)
Satisfaction/ turnover	43	Wikipedia as community (6) Wikipedia bureaucracy (5) Opinion about Wikipedia (8) Would change Wikipedia (9) Circumstances / reasons for leaving (15)

Following this phase I started to think of ways to interpret patterns of participation in light of my interview data and of relevant sociological literature, by tying together existing and emerging theories (Glaser and Strauss 1967; Strauss and Corbin 1998). Theoretical questions led me to refine the *contributor role* category into two different types of participation, *topic-focused* and *broad* participation. This latter differentiation emerged as participants suggested that social closure effects are manifest in situations of repeated collaboration (Coleman 1990). Last, the final coding categories expanded over two main levels, attending to *individual agency*

(*individuals' acts* and *relationships*), and *structural context* for the interviewees' actions and interactions.

This latter classification builds on both Hackman's and Strauss and Corbin's observations that multiple levels are necessary and interdependent in the quest for sociological explanations (Hackman 2003; Strauss and Corbin 1998), because minimizing or leaving out socio-structural conditions in which actions and interactions take place deprives the explanation of its contextual complexity and risks to oversimplify mechanisms. In attending to both levels, my analysis attempts to consider the effects of both structural conditions and individual agency on the phenomenon of interest. As further discussed in this study, these levels interact and mutually influence each other: participants shape the interaction environment while their environment shapes their actions and interactions.

By analyzing the relationship between individuals' structural position measured as dependence on their social network and their participation patterns, I propose a social mechanism which may account for the observed turnover patterns. This mechanism showcases a process through which embeddedness in collaborative work, under conditions of high dependence, can have a negative impact on individual commitment. In the next section I characterize the turnover patterns observed in the data and examine their relation to initial participation, intensity of past participation, choice of role in Wikipedia collaboration work, and social networks of co-contributors and non-contributors (offline relationships).

FINDINGS

In this section I first describe the study findings in terms of variation in interviewees' decision to leave Wikipedia as a function of their task-related activities and social relationships in order to identify the antecedents of turnover in Wikipedia. I then delve deeper into the

identified factors to unveil the social mechanism underlying the link from the antecedents to the turnover decision.

Since the outcome of interest is organizational commitment, measured as continued participation in Wikipedia, I start by identifying the three main categories of respondents: stayers, peripherals, and leavers (see **Table 3.2**). While peripheral contributors clearly differed in contribution patterns from highly active contributors, I consider them together as "stayers" in the analyses given that "peripherals" have suggested that their reduction in the frequency of contributions to Wikipedia was temporary and exogenous to their relationships and activities online.

Table 3.2.   Dependent Variable Description

Contributor participation	# Interviewees	Description
Highly active (stayers, S)	13	Continues to contribute almost daily
		Expresses desire to continue contributing
Occasional (peripherals, P)	11	Contributes less than he used to
		Expresses desire to continue contributing but cites exogenous causes why contributes less
Departed (leavers, L)	11	Unequivocally left Wikipedia
		States no intention to contribute again / return

At the onset of this study I considered the possibility that stayers and leavers may differ in their initial entry in the Wikipedia environment, in the extent to which they have engaged in contributions at the peak of their participation, in the extent to which they have experienced disappointment with regards to collaboration norms, in the types of tasks and participation roles that they self-select into, or in the importance or structural characteristics of their online and offline networks. In the following sections I compare and contrast qualitative evidence from interviews with stayers and peripherals with information from leavers, in order to identify the antecedents of turnover in collaborative production.

Initial Involvement

I started by examining stayers' and leavers' initial participation patterns, in order to observe

whether there were differences in their starting motivation for Wikipedia contributions. Most

interviewees have shared that they have started participating in Wikipedia article writing

gradually, by first making minor corrections or low-cost contributions to articles related to their

hometown, occupation (L5, P8 and P9)[78] or hobbies (L6 and S8). Most interviewees described

the editing process as easy to learn and engaging:

> I got onto one article [where] I knew the subject matter quite well and I saw an error. So I clicked on the 'Edit' tab and [discovered that] it was remarkably easy to make changes. (S3)

> I edited an existing article and … after spending a couple of days … following what changes people made to it, [I started] editing other articles and following changes to them as well. (P7)

> My first contributions were copy-edits to articles that were marked in need of copy-edits. I found it pretty simple to contribute to, and I became more involved soon afterwards. (L2)

> I started with small edits, expanding topics that I knew something about from personal experience. But then I moved to rewrite all the articles on the video game industry. (L6)

> Since the initial conditions of participation by stayers and leavers were similar, I turned

to examining the difference in their subsequent participation patterns.

Intensity of Past Involvement

Wikipedia contributors interviewed spent anywhere from a few minutes to a few hours a week

making changes to articles and engaging in conversations with other contributors. For example

L7 had periods when he was contributing about nine hours a day, while S8 continues to

contribute about 20-25 hours a week. Most interviewees suggested that way Wikipedia is

---

[78] From this point on, I will be referring to the interviewees using designation of category (stayer, peripheral or leaver) and the number of the interviewee in my records, such as S5, P4, or L5.

organized makes it easy to discover new opportunities to contribute and to enter conversations

with other contributors and to stay engaged:

> Wikipedia is so easy to fall into because it's so useful. Suppose I'm reading a [news] article and I come across words I don't know [so] I might look them up on Wikipedia. And because there are so many links on the article I will start tumbling through the world of Wikipedia. Everything is so deeply interconnected that it's just very easy to wander through ... I'm reading [an article] and … I don't know what that [term means], and I'll click another link and it'll take me to a new page with new ideas, people …it's a virtuous or vicious circle depending on your point of view. (P1)

As the quote above suggests, editing on a variety of topics with other contributors is draw

for participants. Individuals learn new information by exploring Wikipedia, and gradually

discover more articles of interest and potential collaborators. Contributing to article writing is

not, however, without pitfalls. Given how the wiki platform is set up, there is in fact an

unavoidable trade-off between individual and collective benefits from contributions. I turn to

examine whether leavers were individuals particularly disappointed with Wikipedia

collaboration norms and their enforcement.

The Social Structure of Collaboration

The trade-off between individual and collective benefits from one's contributions is inherent to

collaborative work. On one hand, the collective clearly benefits from including only the best of

work in the common product. On the other hand, individual contributors may get upset if they

feel their efforts are underappreciated, or rejected for unsubstantiated reasons. The decision of

rejecting some contributions is therefore fraught with social costs. For the contributor whose

effort is being judged, high costs are incurred if he feels wronged, especially regarding a

substantial contribution. For the norm enforcer, costs come from disapproval by the sanctioned

member or by third parties if the norm or evaluation is not shared. For this reason, many

participants may prefer benefiting from norms without contributing to their enforcement (Baron, Kerr, and Miller 1992; Coleman 1990), which could lead to loss of article quality.

Wikipedia is a place where this trade-off is particularly acute for two reasons. First, contributions are asynchronous, in the sense that contributors make article changes and their co-participants react to these. Hence contributions may be rejected after participants have put in substantial effort, so they see not only their ideas but also the value of their work rebuffed. One contributor (L11) explained: "I did my graduate work on Asian music and have written [Wikipedia] articles on traditional musical instruments from Korea. I put a complete list of these instruments on Wikipedia, but this Korean-American high-school kid has been introducing incorrect information to the page. I pointed this out about a hundred times on the discussion page, but this person keeps getting other [editors] to overrule me. [Sometimes]… on Wikipedia, it's not always the most informed person that wins, but it's the most persistent." Second, those whose contributions are rejected are a higher risk of misinterpreting this rejection in absence of a rich communication medium. If the rejecting party does not clearly and respectfully communicate the reasons for rejecting a contribution to the affected party, the affected party may fail to learn how to improve on his work and may feel that his contributions are not useful or appreciated. It is important that individuals who are motivated to invest effort in contributing are assured that their edits are fairly judged and stand a chance to become part of the article. Otherwise, they may leave because they feel their work is not sufficiently appreciated or respected, or because they are tired of frequent conflicts with other editors over the merits of their work.

If interviewees were to systematically differ in terms of articles they are contributing to, we may expect that those contributing to highly controversial articles to leave Wikipedia faster

123

than those participating in the writing of less controversial articles. However in my interviews I found that respondents described a similar range of situations encountered, and that all of them had enough knowledge about Wikipedia as a body of articles such that they could potentially self-select into a type of collaborative environment that they were comfortable with. This finding may be attributable to the fact that most respondents had large numbers of annual contributions to Wikipedia, varying from about 110 (P5) to 32 thousand (S7), with a mean of about five thousand contributions and a standard deviation of about 6,250 edits. All interviewees had similar opportunities to socialize into Wikipedia, understand its norms and interact with other contributors. Many of them have noted that they experienced both productive collaboration and disagreements with fellow contributors, and that sometimes they had to negotiate compromises which led to article changes away from their original intent. However, some interviewees left and others continued contributing as a response to similar situations. This brings us back to the central puzzle of this study: what is different about individuals who remain committed to the organization and continue contributing despite encountering similar conflicts and norm infringements?

In my analysis I found that a good understanding of contribution norms and collaboration guidelines is a necessary, but not sufficient, condition for editors to remain committed to participating in Wikipedia. Both stayers and leavers listed similar shortcomings of Wikipedia. For example, both interviewee L1 and S8 complained about edits by anonymous contributors and suggested that registration should be mandatory for everyone who contributes, presumably because the intentions, skills and reputation of a registered contributor can be more readily assessed from contribution history, whereas one cannot retrieve a track record for anonymous contributors.

Contributor Roles

The concept of contributor roles has emerged naturally from stories about participation patterns. Several interviewees suggested that establishing one's domain of competence and building a reputation for a specific type of contributions such as copy-editing or vandalism removal, or on a narrow knowledge topic raises the odds that one's work will survive collaborative editing. For example, contributor P11 shared: "[Another editor] knows how to make diagrams for chemical structures. Sometimes I see an article on a chemical compound that does not have a molecular diagram, so then I send him a quick message, and usually within 24 hours there it is." S5 explained: "My articles focus on articles about culture, musicians and music education in Serbia. Our musicians deserve to be remembered, and currently you can only read about them [only offline]. I got into Wikipedia by starting my own articles and then improving them. This is by and large the only thing I do on Wikipedia." P8 shared: "I started with the harbor of [N] page, and spent a lot of time on that page, because of my business, [N] Harbor Tours. But from there I've expanded and I've worked on many different pages; things relating to other parts of [N] and then moved to other things." A similar strategy is echoed in S3's story:

> The interesting thing about Wikipedia is that people can check you out very easily, very quickly, and find out some basic information about you and what you've done on Wikipedia. So you have a certain status … that people [can] recognize…

> Very early on I developed kind of my own approach to articles. And it sort of played to my own field and strengths. I've worked as an editor for publications [offline]. So it was natural for me to do a certain kind of editing. So I found my [place] in the Wikipedia community. If I copy-edit an article, it's not likely to get [undone]. Nobody is going to come on and [argue]. I think they have a certain level of respect for what I do. (S3)

> Interviewee P5 adopted a similar strategy, because it was aligned with his offline

commitment: "I felt I could correct things that weren't right ... Part of my experience includes

work as a tutor, [and my participation is] an extension of that." Both P5 and S3 adopted a strategy of specialization in their edit types. One major difference between their participation patterns however turned out to be that S3 deliberately selected specific topics to contribute to:

> I'm Canadian, so I looked at several Canadian articles and I saw systemic biases in them. They were written obviously by white, techno-literate young males mostly…So I set as my goal to start working on Canadian articles that input those groups. So for the next several months I just worked quietly. (S3)

In contrast to S3, who as of January 1, 2009 made about 30 contributions per day in the previous month, interviewee P5 had made only five edits during the previous two months. He shared that he had applied his skills to a broad range of topics and was disappointed with most subject-specific participation: "There's a 'ruling clique' … in most subject areas. They have their status quo and are loathe to embrace anything [else] …I don't like the community there, it's a 'survival of the most persistent' where factual accuracy is usually but not always the victor." While some contributors adopted the broad contribution strategy, the topic-focused approach has been adopted by interviewees who specialized in articles on (usually) one specific subject, such as music and musicians (S5 and S7), biology (P8 and P9), tropical cyclones (L4), weapons (L6) or comic-books (L3). This pattern of contribution has the advantage that one may edit more frequently together with a core group of recurrent contributors, which provides social closure, a structural property that facilitates norm enforcement and social rewards provision to participants (Coleman 1990).

While several interviewees have expressed disappointment similar to P5, stating that broad contributions exposed one to the risk of being under-valued in topic-centric discussions, I found that this dimension was inconclusive in predicting the likelihood of contributor turnover. In order to assess the benefits of social rewards from co-participants I turn to examining the issue of social interactions in Wikipedia collaboration structures.

Social Networks with Co-Contributors

In Wikipedia, frequent participants are referred to as 'Wikipedians.' Surprisingly, I found that despite the high levels of participation of interviewees, many explicitly reject the idea of self-identifying as 'Wikipedians'. P1 explained: "I don't think I've ever quite fully identified with the project. I joined it, but I could see that there were other people who were more obsessed with it that I was. … I [don't] see myself as a Wikipedian". S5, a participant with about 2,000 contributions, confessed a similar feeling, indicating that he welcomes communication with fellow contributors but does not think of the participants engaged in collective production on Wikipedia as community: "I communicate with [other editors] occasionally, but, generally, I don't feel a sense of community between Wikipedia editors. I prefer a sense of community between people in the real world." Leavers (departed contributors) expressed similar reluctance to identify as Wikipedians, but they were more likely to state that they felt allegiance towards particular individuals or Wikipedia subgroups such as the group of contributors writing about comic books, warfare or tropical cyclones.

My analysis revealed that not only do the stayers and peripherals reject identification with the community at large, but they also tend to form weak ties with other contributors (Granovetter 1973) and restrict communication with others to task-oriented communication, related to the article or topic they are working on. This pattern largely manifested by stayer and peripheral participants was coded as a *low social participation*:

> I've only had one conversation on Wikipedia, and that was when I was very concerned about what they call neutral point of view. And someone had listed one of my favorite publications, *The New York Times,* as a liberal publication. And it dawned on me that … the world liberal was far too vague and *The New York Times* far too complex ...So I thought … I'll remove it. And I came back the next day and it was back. [So we started changing the article back and forth which was annoying to both of us.] And I decided this wasn't a good idea and that I would actually talk this through. (P1)

The low social participation of stayers and peripherals is not due to their lacking information about all the ways in which they could connect with other participants. As P1 above illustrates, they would talk to others if there is a need for communication, to clarify a misunderstanding or sort out a disagreement regarding an article. As S3 states below, stayers and peripherals may even acknowledge that it may be important for others to get to know their co-participants better, or "face-to-face," but they generally focus on "factual contributions" rather than social interactions:

> There are a lot of [contributor] meet-ups and they're big events... There was one recently for people in British Columbia… and I was tempted to go to that, because there are some interesting people that I've dealt with [on the project], but in the end I didn't. It's important, I would say, for many people within the Wikipedia community to get to know, face-to-face, their collaborators. It hasn't been that way for me… my contribution is more focused around factual contributions and not so much the interaction with others. (S3)

In contrast, leavers have a history of being socially involved with other participants. One leaver has shared the story of collaborating with others in articles on his topics or interest:

> A few people, notably John, and Terry, as well as James[79] were quite helpful in the areas I was contributing to - Hawaii, military stuff, and firearms. [W]e were able to accomplish a great deal of work by combining and distributing workloads. (L6)

The leavers reported not only more engagement in coordinating contributions with others, but also social interactions with fellow contributors, which suggests that they developed multiplex relationships with Wikipedia co-participants. One leaver (L5) reports asking for "help on non-related personal matters", and another (L1) reports that she frequently chatted with other Wikipedia participants and even met her "best friend" in this setting:[80]

---

[79] Contributor names were altered for privacy reasons.

[80] Typically, multiplex relationships have a tendency to be strong ties (Brass, Butterfield and Skaggs 1998).

I would occasionally use the talk page function a bit like an internet chat, and in addition used the e-mail function to get help on non-related personal matters such as travel advice or similar, based on knowledge obtained or inferred through interaction with other wikipedians or from personal information placed on individual user pages by the users themselves. (L5)

I use IRC and chat with some editors on Yahoo messenger. I've made some amazing friends, including the one I refer to above as my best friend... I got to know some editors from interacting with them on Wikipedia and becoming friends with them through off-wiki chat. [I met my best friend through policy discussions, and then] we joked around on our talk pages. Eventually, a mutual friend shared our Yahoo IDs, and we ended up talking in Yahoo. Craziness ensued. Haha. I met an amazing person who is so much like me... We had a rollercoaster of a friendship, wrote an article together, created a project together, [and] caused some drama together. (L1)

My analysis suggests that, contrary to the expectation that social embeddedness benefits participants through higher emotional engagement in the organization (Kahn 1998), highly embedded participants like L1 and L5 are the ones least likely to remain engaged.[81] Even more significantly, contributors whose social networks with other contributors are of multiplex nature (Burt 1983; Verbrugge 1979) whereby the focal actor uses the relationships not only for task-related purposes but also for friendship and advice are the ones most likely to leave Wikipedia. This is contrary to what we would expect based on an escalation of commitment logic: the more socially embedded contributors are, the more we would expect them to be led towards persistent commitment by psychological and social forces (Staw and Ross 1989). To explain this paradoxical finding I turn to examining the importance of social networks in Wikipedia relative to other social ties in interviewees' lives.

---

[81] In order to address concerns that the quality of stayers' contributions may be different (substantially higher) than that of leavers' contributions, I examined the likelihood that each experiences erased contributions (reverts). I found no clear-cut evidence of a difference between the two. Examining the survival times of leavers' and stayers' contributions would be a more effective measure of work quality, but this measure cannot be obtained since my quantitative Wikipedia datasets to not include the actual contribution texts.

Social Networks with Non-Contributors

The data suggests that the collaboration structures on Wikipedia cannot be evaluated without a proper understanding of their relative importance to the focal actor. I find that contributors whose real life social alternatives are limited or otherwise unsatisfactory place a much higher weight on receiving intrinsic and social rewards when contributing on Wikipedia. If individuals have little minimal offline social networks or other social involvement (are under-involved offline), and if the value of their Wikipedia contributions is questioned, their disappointment is more salient than editors' who can turn to offline social alternatives for support and rewards. When these former editors' expectation of social rewards from collaborators is not met, or when conflict arises, they are more likely to abandon the project because they fail to distance themselves from that event.

In the analysis I estimated interviewee's offline social involvement and networks from information that interviewees provided about their education level, occupation, marital status, and friendship networks. Responses varied from the under-involved former contributor L1 to the socially active S8. The former explained that her real life does not provide much social support and intellectual satisfaction:

> I am a waitress. I am married but with established plans to separate and divorce …I do talk [with my friends] about [Wikipedia], but they do not contribute. I told one [about my contributions], and she responded, 'Is that the book that tells you the definition of words?' Sometimes I think it's best they don't contribute. (L1)

In contrast, S8, a California engineer who contributed over ten thousand edits, explained his real life social network as follows:

> I probably have 200 friends of various kinds… I have a few that are so close [that] I would take a bullet for them or something… maybe six, and then so many more that I enjoy talking with. [Some] have different career connections with me, some have no career connections with me, but a lot of my current friends are ones that I've met while I

was working, and very few are from before I was working. [V]ery few are from school, but some are… I'm certainly no hermit. (S8)

S8's intrinsic motivations to contribute may be similar to everyone else's, but when his edits are rejected or altered, he sees this interaction as inherent in Wikipedia's collaboration format and looks at Wikipedia from the perspective of its mission to provide knowledge to the world. Nevertheless, S8 voluntarily mentions his difficulty in collaborating with some editors who are inflexible in their Wikipedia actions and lack a sense of perspective regarding their contributions:

> Some people have nothing but Wikipedia in their lives, and … maybe … almost no social interaction, almost no family interaction. … A few are so tied to the rules of Wikipedia that they don't see how the common sense of an information source like Wikipedia would best be served. (S8)

Up to this point, the analysis suggests that participants with strong offline social networks, whose family and friends provide social rewards and an alternative to Wikipedia participation, are more likely to continue contributing compared against contributors who are more dependent on Wikipedia for social relationships and rewards. As S7's story below portrays, stayers and peripherals are able to distance themselves from conflict and accept it as inherent part of collaborative work, and, in general, social interaction:

> [Wikipedia is] just like any small group or large system. It has the same dynamics. [Discussions are] very heated, a lot of backstabbing, people trying to get their friends to get involved. … [Sometimes] you can't get any justice. In a way it's like society at large. When you appeal to a higher authority, sometimes you get ignored. [It can be irritating but] I don't think it's a flaw of [Wikipedia], it's just a flaw of social systems that shows up on Wikipedia just like it might show up [anywhere else] in society. (S7)

In a nutshell, S7 recognizes that social interaction on Wikipedia is similar to any other social system, and is not upset over Wikipedia conflicts and "backstabbing" he perceives as inherent to collective production. S7 is an American male graduate student, with a small but

131

dense personal network offline: "Out of those six close friends that I have, four of them are friends amongst themselves, and the fifth one knows them but is not as close to the others. And the other one is my good friend from college back home, and I don't see him that often." Like many other interviewees, he doesn't often mention his work on Wikipedia to his offline social network. Some of his close friends may know he is contributing, but "Certain people I withhold information from, I don't want them to know that I do it. So for example, my dissertation advisor doesn't know...I hope… [If I tell people that I contributed over 1,000 edits] instead of saying, 'wow, that's great', [I get] a quizzical reaction. Like, why would you devote yourself to doing something that's wasting your time like that … you're not getting paid. [O]ftentimes people are half-impressed but also a little weirded out."

Although substantial, S7's Wikipedia involvement described above represents only a small portion of his overall identity and social interactions. S7 is a fairly representative case: stayers and peripherals are characterized by the lower social participation in Wikipedia and high social involvement in their offline lives. **Table 3.3** presents a summary of interviewee attributes as they pertain to social participation online and offline, year when the interviewee started contributing, percentage of contributions deleted, number of average contributions per day, age of the interviewee (at the time of the interview), and highest educational level attained or in progress. In this table, a relationship (marriage, partnership) and/or offline friends was coded as '1' and none of the two -'0'; having Wikipedia friends one engaged in social exchanges with was coded as '1', and not having such friends as '0'; and a high school education or equivalent was coded as '0', college education as '1' and graduate level education as '2.'[82]

---

[82] The offline social network variable was dichotomized because interviewees provided several types of committed relationships as response, and because several declined to answer the question, while providing a wide range of offline network estimates as well, from "about 200 people I am in contact with on a regular basis" to "six people that I would take a bullet for" which made it difficult to normalize the size of the networks across the sample. Most

Table 3.3. Summary of Relevant Interviewee Characteristics (N=35)

Participation	Education	Age	Year Start	Edits / Day	Deleted%	Soc. Offline	Soc. online
Stayer	2	27	2006	1.8	4.85	0	0
Stayer	2	37	2005	89.2	1.69	1	0
Stayer	0	50	2001	5.8	1.99	1	0
Stayer	1	26	2001	36	8.84	1	0
Stayer	2	54	2003	8.4	1.96	1	0
Stayer	1	21	2007	24.7	14.29	1	1
Stayer	1	22	2005	16.8	2.63	1	1
Stayer	1	16	2008	0.5	8.34	1	0
Stayer	2	69	2006	8.1	1.87	1	0
Stayer	1	.	2006	17.6	6.96	1	1
Stayer	0	29	2003	1.8	5.53	1	0
Stayer	1	70	2007	1.4	2.25	1	0
Stayer	1	.	2007	44.7	1.04	1	0
Peripheral	2	41	2006	1.7	3.93	0	0
Peripheral	1	23	2007	13	3.25	1	1
Peripheral	2	47	2008	8	5.7	1	0
Peripheral	2	33	2002	1.8	3.32	1	0
Peripheral	1	40	2001	0.3	2.26	1	0
Peripheral	2	27	2005	4.6	13.25	1	0
Peripheral	1	40	2002	15.7	2.74	1	0
Peripheral	1	19	2006	0.2	4	1	0
Peripheral	1	35	2004	9.8	4.94	1	1
Peripheral	2	32	2002	3	2.68	1	0
Peripheral	1	21	2004	2.1	4.42	1	0
Leaver	2	27	2005	9.8	8.61	0	0
Leaver	2	50	2005	11	5.11	0	0
Leaver	0	26	2006	22.5	10.55	0	1
Leaver	1	.	2004	17.8	43.07	0	1
Leaver	0	19	2005	31.3	6.43	1	1
Leaver	0	18	2005	4.5	4.31	1	1
Leaver	0	30	2004	2.6	16.99	1	1
Leaver	1	23	2004	14.7	9.16	1	1
Leaver	2	31	2005	11.3	1.27	1	0
Leaver	1	52	2005	12.5	13.52	0	1
Leaver	1	.	2005	31.4	3.61	1	1

interviewees provided no / tentative estimates of their social network density – this lead to dropping this variable from the analysis. In terms of Wikipedia-based relationships, many interviewees proceeded to list people or describe events related to interactions with online friends; for analysis purposes I coded mentions of such social interactions (beyond pure coordination in article writing) as "1" and explicit denial of online friendships as "0."

A brief ordinal logit regression of the data[83] in **Table 3.3** using participation as dependent variable provides support for the qualitative findings.[84] Out of eight variables included in the specification, only social participation online and social participation offline attained statistical significance, indicating that for a one unit increase in one's offline social network we expect a 2.03 increase in the log-odds of continuing to participate (90% confidence level) while for a one-unit increase in one's online social participation to Wikipedia we expect a 2.59 decrease in the log-odds of continuing to participate to Wikipedia (95% confidence level).

In short, the quantitative analysis described provides support for the assertion that individuals who are highly dependent on Wikipedia collaborators for their social interactions are the more likely to exit, compared against individuals whose offline social networks provide a buffer against inherent disappointments and conflict associated with collective production. If, following Emerson (1962) we relationally define dyadic dependence as *"of actor A upon actor B as (1) directly proportional to A's *motivational investment* in goals mediated by B, and (2) inversely proportional to the *availability* of those goals to A outside of the A-B relation,"* we can see that stayer and peripheral dependence on Wikipedia is low, whereas leavers' stories hint to a high dependence on their Wikipedia network.

DISCUSSION

*Linking Power-Dependence to Turnover*

---

[83] Interviewee identification codes were excluded to preserve the confidentiality of information, since one may be able to identify individuals by triangulating interview and table data.

[84] The results of this ordinal regression are available upon request. For analysis purposes, education was coded as separate dummies for college and graduate level (with baseline = high school education). Out of all interviewees, 10 were students, 20 fully employed, one unemployed and 3 retired (one declined to answer). The 'work status' variable was not included because student / employed status was not significantly correlated with participation. The adjusted R-squared value suggests that this specification explains 23.25 percent of the variation in participation patterns; a simple regression with 2 degrees of freedom including only social participation online and offline explains approximately 15.5 of the variation in participation patterns. The Brant test for the proportional odds assumption confirms that the parallel regression assumption was not violated.

Due to their low dependence on their Wikipedia social network, stayers and peripherals perceptions of self-efficacy and their engagement in Wikipedia are not affected by the conflicts inherent in collective article writing. Conversely, leavers were highly dependent on their group of Wikipedia collaborators both with regards to task-related outcomes and personal outcomes such as receiving status, advice or even friendship from others. In his theory of conflict, Simmel (1964:44-45) suggests that this high dependence can engender pernicious conflicts because "a quarrel arises between persons in [an important] relationship, it is often so passionately expansive [because we invested] with the totality of [our] being and feeling" in that relationship.

Consistent with Emerson's (1962) power-dependence theory we therefore observe that contributors highly dependent on Wikipedia engage in *balancing operations,* "structural changes in power-dependence relations which tend to reduce power advantage." In this case, we observe actors choosing withdrawal; "the denial of dependency involved in this balancing operation will have the effect of moving actors away from relations which are unbalanced to their disadvantage" which effectively leads to turnover in the context of participation to Wikipedia.
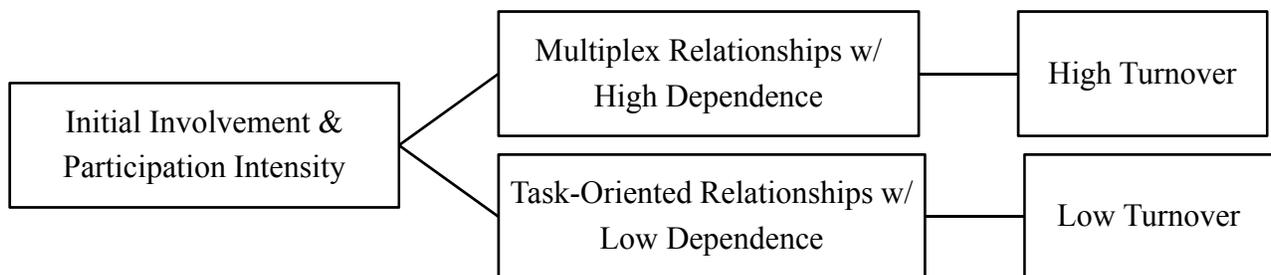
Figure 3.1. Participation Patterns and Turnover Process

The findings in this study point towards a divergent process model (Van de Ven 1992) whereby individuals undergo a developmental sequence, from becoming involved and socialized in a similar manner in collaborative production settings, to choosing different roles and strategies

for building relationships within the organization, and to divergent outcomes in terms of commitment to the organization (see **Figure 3.1** above).

Patterns of the Departed

In my analysis I identified a difference in the social structures of contributors who are still committed to editing Wikipedia and those who explicitly abandoned the project. There are many reasons why people would stop editing online. Exogenous factors such as a more consuming job, family demands, or traveling impinge upon online participation. Several peripheral interviewees offered such reasons for their reduced edit volume. However, there were some who stood out: those who departed bitterly, after a conflict or major disappointment. For example, L5, a former contributor who focused on cartography articles "would occasionally use the talk page function a bit like an internet chat … to get help on non-related personal matters such as travel advice or similar, based on knowledge" from interaction with other contributors. His edits evolved from work on cartography topics to talk page discussions and socializing with other contributors. He eventually stopped editing completely stating that he "came to view the editing process as arbitrary and random" (L5) and that he was disappointed that people he collaborated with did not respect his work.

Instead of building a task-related network of collaboration, L5 shifted towards more intensive socializing, forsaking the social closure and capital generated in the act of actively collaborating with enthusiastic others on a shared goal. Another departed editor (L1) shared a similar story: she developed a close friendship while contributing and, when her friend ended up being banned from the project for disrespectful behavior, she ceased editing; her motivation to participate was dramatically affected when he left. This decision was not necessarily caused by a socio-emotional orientation to the project – in contrast to stayers' and peripherals' more

136

instrumental one; as suggested by the analysis, neither group identified extensively with Wikipedia. The difference in participation behavior can be instead traced back to the formation of different types of ties with other contributors.

These findings are consistent with research examining the relationship between coworker support and voluntary turnover. Despite the fact that the organizational support theory literature (Rhoades and Eisenberger 2002) has proposed that perceived support in the organizational environment is associated with a reduced likelihood of turnover, studies investigating the relationship between perceived coworker support and turnover failed to find evidence of this effect (Iverson 1999; Iverson and Pullman 2000). In a similar vein, network theorists have pointed out that there is a downside to high embeddedness. A qualitative study of work relations in a garment district exposed a "paradox of embeddedness" of inter-firm ties in social relations: in highly embedded networks, feelings of obligation, friendship, or betrayal in interpersonal relations may be so intense that emotions override task-oriented imperatives (Uzzi 1997). Work by Portes and Sensenbrenner (1993) indicates that a similar phenomenon occurs in ethnic enclaves, where immigrant businesses are highly dependent on co-nationals both for clientele and labor market supply.

Patterns of the Stayers

In contrast to editors who have departed, interviewees who are still participating to the project told a story of respectful task-oriented interaction with fellow contributors. For example, in the quote below S11 explains that he has minimal information about the identity of fellow co-editors but is familiar with their skills and areas of knowledge. S11 openly engages potential collaborators in discussions relevant to the task, and is respectful of their expertise but does not attempt to socialize with co-contributors:

Out of the ten editors whose collaboration and expertise I really value, I don't really know much about any... I don't know where they live, I don't know their ages. I may know something about their cultural backgrounds. For example.., I think this guy lives in Canada somewhere and he's like a Chinese-Canadian. And he's really knowledgeable about Chinese esoteric ingredients, particularly … sauces and things…so if I'm interested in something and I don't know about it, he can read Chinese, so then I'll say… 'I'm working on this article - can you look [something] up?' (S11)

S8 developed a pro-active strategy of collective production: to let people express objections before he puts too much effort into editing, he preemptively announces his planned contributions and stays open to fine-tuning details of his work: "I said, "Here's what I planned," and they said, 'Good. If you need any help …put another message here' … I didn't feel like I was a part of their project [but] I was doing stuff that affected their project. [M]ost of my [effort] goes to articles," not social interactions. When contradicted, S8 considers the context of the argument and other parties' credentials, and accepts compromise as inherent in collaborative work.

In brief, the interview data suggest that stayers and peripherals are highly active in their offline lives, and walk a narrow line between socializing too much and too little on Wikipedia. Peripherals often contribute less because of exogenous reasons - as S4 explained: "I was going to do my master's immediately after a college, but I found that I was learning more in university libraries writing Wikipedia articles than when I was a student. So I spent the next full year contributing on a full-time basis, working at libraries, researching things, introducing new pages…When I got my first job, I kept editing a fair bit... Nowadays I spend a little bit of time each day on Wikipedia, probably less than an hour." S1 explained his fluctuating participation: "Now I'm roughly around like 1,100 edits a month. Last month I was a little bit lower and that was because I had more work for my classes and I was concentrating on finding a new apartment

and I was doing my application for study abroad. Basically just real-life concerns. Real-life things will obviously take up time from Wikipedia."

Too much socializing, as in the case of L1 and L5, leads to losing focus on the article writing task, and to lost motivation to participate when friends are not around to socialize with. Stayers' social participation and identification with Wikipedia are minimal, but they are open and focused on communicating about their topics of interest which leads them to be recognized as experts in their particular areas.

CONCLUSIONS

This study proposes a social mechanism through which individuals' commitment to participation in an organization is affected by the social structure around their collaborative work and by their social involvement with others outside the organization. Qualitative research based on archival data and interviews with participants to Wikipedia suggested that individuals who restrict their social participation to weak ties with collaborators are likelier to continue participating. Additionally, participants whose social networks outside the collaborative production setting provide social rewards are less likely to stop contributing because their dependence on the social network of collaboration is low.

The analysis of online and offline social networks, and, even more generally, of the connection between the relative importance of organizational networks in the structure of an individual's social life and conflict among organizational members represents a novel area of exploration for the field of organizational behavior because it integrates and extends the sociological literatures on conflict, social networks, and power-dependence. Additionally, the finding that over-involvement in social interactions within organizations exposes individuals to higher risk of turnover and distracts from task-oriented interactions suggests that organizations

139

should consider the cost of drawbacks of fostering strong ties among members at the expense of life-work balance.

This exploratory analysis of turnover in an online voluntary project work opens the door for a wider array of social studies looking at the mechanisms through which social structures affect participant commitment. Several limitations of the current study recommend this area as a fertile ground for research. First, this research is largely based on interview data from an online volunteer setting. While one would expect that findings from this setting are not directly applicable to work in traditional workplaces, I argue that role conflicts may even be less salient in volunteer settings since participants in volunteer work are known to have more children, be more socially active and be more engaged in religious activity than non-participants (Wilson and Musick 1997). Another limitation is that lack of individual commitment in the workplace may not immediately translate in turnover, due to the structure of rewards in these settings. Therefore examining the relationship between turnover and individual dependence on the organizational social network in a traditional workplace setting, under conditions of collaboration similar to those in Wikipedia would be a fruitful next step to extend this research.

Second, this qualitative analysis fosters theory generation in linking organizational network dependence and turnover; further research is needed to test its applicability. The purpose of this paper is to identify a social mechanism responsible for individual turnover in collaborative production settings. Examining this mechanism in the context of other volunteer and non-volunteer settings, and identifying the necessary conditions for this process to take place are important next steps for generalizing the mechanism.

Workplace engagement and organizational embeddedness are regarded as desirable in contemporary organizations, for their presumably positive effect on individual commitment to

the organization and on reducing turnover. Network structures in the workplace may facilitate provision of social support, thus indirectly moderating individual reactions to negative events such as norm infringement and conflict (Coleman 1990). In an examination of structural, attitudinal, and behavioral predictors of turnover, workplace network centrality, perceived coworker support, and felt obligation have similarly been identified in organizational behavior research as determinants of lower turnover likelihood (Mossholder, Settoon, and Henagan 2005). This latter research has however been largely agnostic to the nature of network ties at work, relying on the assumptions that social embeddedness at work has a positive effect on continued participation, and assuming that there is no interaction between these relationships and participants' overall social network size or structure.

Although "it is frequently recognized that a person's work life needs to be viewed in the context of family and personal concerns" (Kopelman, Greenhaus, and Connolly 1983), rarely did organizational research aside from work-family studies (Bielby and Bielby 1989; Greenhaus and Beutell 1985) integrate these findings in examinations of social network participation at work, or in studies of organizational exit decisions.[85] Managers often assume that multiple commitments and roles are antecedents of turnover because roles expand one's repertoire of behaviors and set of obligations, which may lead to conflicts (Goffman 1959), and therefore foster relationship formation in organizations without regard, or even to the detriment, of other relationships. The present study suggests that without robust evidence to support these assumptions these practices may lead to dependence on social networks within the organization and ultimately result in individual turnover.

---

[85] Organizational commitment has been "generally overlooked as a variable of interest in work-nonwork research" (Lounsbury and Hoopes 1986) as well, which points to the need for research on the relative importance of organizational networks in the context of individuals' social life for turnover decisions.

APPENDIX A. VANDALISM, UNDO, AND REVERT OF UNDO EXPLAINED.

Vandalism

We first sought to discriminate undos of vandalism, which are legitimate actions, from undos of good edits, which are counternormative. To identify undos of vandalism, our sorted all the versions of a given article chronologically and analyzed each version starting with the oldest (see **Table A.1** below for a sample article). The algorithm relied on the fact that an undo of vandalism creates a version of an article identical to a previous version. Because it would be too time-consuming to compare each version to all previous versions, we relied on the simple shortcut of comparing the lengths of successive versions.[86] That is, if no prior version of identical length existed, we concluded that the version in question could not be an undo of vandalism.[87] But if the algorithm found a previous version of the same article with an identical length, it examined all versions between the one in question and the previous version of the same length.[88] The algorithm then tested whether intermediate versions by the same editor were less than 10 percent of the size of the version in question. If so, we recoded all intervening edits as acts of vandalism and coded the version in question as an undo of vandalism.[89] In examining version 263 in **Table**

---

[86] Another option would be to identify an undo of vandalism by examining the short notes that editors sometimes append when undoing an article version. But these notes are optional, and therefore unreliable.

[87] In some cases acts of vandalism take the form of very small changes to an article, such as inserting a vulgarity into the text. When the next editor undoes such an addition, this is technically an undo of vandalism. But because our algorithm identifies only large changes to articles as vandalism, such an act will be coded as an undone edit followed by an undo. The algorithm will thus underestimate the rate of vandalism and overestimate the rate of undone edits.

[88] If many versions of the same length were found, the algorithm would select only the most recent. For example, if the algorithm was currently analyzing version 263, both versions 242 and 261 would be identified as the same length as 263. The algorithm would then select version 261.

[89] It is conceivable that an act of vandalism that removed more than 90 percent of an article's content was followed by one or more versions and then by an undo that restored the article to its original state. To take this possibility into account, we assumed that the currently analyzed version was an undo of vandalism even if only one version existed with less than 90 percent of its content. We also tried coding the original edit that had removed 90 percent of article content as vandalism, the subsequent versions as regular edits, and the undo as an undo of vandalism. Given the rarity of such editing patterns, all coding schemes resulted in the same pattern of results.

2.7, for example, the algorithm would discover that version 261 was the same length, and that

version 262 was less than 90 percent as long. We would then code version 262 as vandalism and

version 263 as an undo of vandalism.

Table A.1. Sample Article History

Version	Date and time	Editor or IP	Text length in bytes	Final designation of edit type[90]	By
…	…	…	…	…	
242			94,399	Regular edit	
…	…	…	…	…	
261	Feb 4, 2005, 23:20	AndrewP	94,399	Regular edit	
262	Feb 4, 2005, 23:25	PettyCrime	10	Vandalism	
263	Feb 4, 2005, 23:27	DogEatDog	94,399	Undo of vandalism	
264	Feb 4, 2005, 23:15	DannyP	94,134	Regular edit	
265	Feb 5, 2005, 12:15	Angela	94,576	Undone edit	Other
266	Feb 5, 2005, 12:15	128.100.91.5	95,333	Undone edit	Other
267	Feb 5, 2005, 12:17	ZZTop	95,134	Undo	Other
268	Feb 4, 2005, 23:16	BriteLite	94,433	Undone edit	Self
269	Feb 4, 2005, 23:19	BriteLite	94,512	Undone edit	Self
270	Feb 5, 2005, 12:10	BriteLite	95,134	Undo	Self
271	Feb 5, 2005, 12:30	ZZTop	95,211	Regular edit	
272	Feb 5, 2005, 1:07	Angela	95,279	Undone edit	Other
273	Feb 5, 2005, 1:09	BlueHawk	95,211	Undo	Other
274	Feb 5, 2005, 1:10	DogEatDog	95,279	Revert of undo	Other
275	Feb 5, 2005, 1:12	Charlie	96,501	Regular edit	
276	Feb 5, 2005, 1:20	MustBeSerious	96,650	Undone edit	Other
277	Feb 5, 2005, 1:25	BlueHawk	96,501	Undo	Other
278	Feb 5, 2005, 1:27	MustBeSerious	96,650	Revert of undo	Self

Undo

We relied on the same logic to identify instances of an undo of a regular edit, which is

normatively prohibited. This action creates a new version identical to a prior version but with no

intervening vandalism edits.[91] If the algorithm detected such a pattern, the version in question

---

[90] Because our algorithm moves forward in time, and we change designations by looking backward, acts initially coded in a particular way may be recoded when subsequent versions of the article are analyzed. Hence, we refer to it as the "final designation"

[91] Our algorithm works only for the original implementation of the undo link, which restored a particular version of an article and removed all intermediate versions between it and the current version. For example, if the current version was 266 and an editor clicked the undo link next to version 264, the platform would create version 267 an exact replica of 264, and discard all changes introduced in versions 265 and 266. In late 2006, Wikipedia changed its software to allow editors to undo changes introduced in only one version while leaving others intact. Thus, in

was categorized as an undo and those between it and the prior version of the same length were designated undone edits.[92] This pattern can be seen in edits 264–267 in **Table 2.7**; version 267 is designated an undo and versions 265 and 266 are both categorized as undone edit.

The algorithm then sought to distinguish between a norm-violating undo and an acceptable one in which an editor created a new version of an article and then, dissatisfied with it, undid his own changes. To do so, the algorithm examined all intermediate versions between the version in question and the previous version of identical length. If all had been by the same editor, it qualified the undo as *undo by self* and undone edits as *undone edits by self*. This pattern can be seen in versions 267–270 in **Table 2.7**. If however the intermediate versions were saved by different editors, all acts of undo and all undone edits were coded as undertaken by other (see versions 264–267).[93]

---

response to clicking the undo link next to version 264, the software would try to create a version 267 identical to version 266 but without changes introduced in version 264. This method retained the changes introduced in version 265. As a consequence, the resulting version was apt to differ in length from the undone version 264, and our algorithm would not work. As a consequence, we limited our analysis to the time period when the original undo regime was in operation.

[92] At this point, it would be possible to compare the texts of the two versions to ascertain that they are actually identical. However, such an exercise would require several terabytes of computer storage capacity and would lengthen the data analysis by many months. We thus use this speedier algorithm. Its biggest shortcoming is that it will assume that two versions are identical when they are merely of identical length, leading us to overstate the frequency of undo and revert-of-undo actions. To test the extent to which this is a problem, we chose 2,000 articles at random and identified every instance when our algorithm found two versions of equal length. We then wrote a short script to test whether the versions tagged as identical were in fact textually identical. Of pairs tagged as identical, 99.7 percent were found to be in fact identical. Such a high rate of correspondence makes us confident that our faster algorithm is relatively error-free.

[93] This algorithm assumed that all undone versions were undone by other, even if some of the intermediate versions were by the same editor who later undid the changes. To verify the robustness of our results, we re-ran the algorithm assuming that every undone version saved by the same editor who later undertook an undo was marked as undone by self and every undone version saved by a different editor was marked as undone by other. Designating undone edits by self and undone edits by other led to coefficient estimates that are in the same direction as those resulting from the algorithm in the main text. We report those from the main text.

Reverts of Undo

Finally, we used similar logic one more time to identify reverts of undo. For a revert of undo to take place, an editor must first undo a prior version of the article, thus making that version an undone edit; another editor then reverts that undo and creates a new version identical to the version marked as undone edit. Thus we can identify a revert of undo by finding that a previous version of the same length has already been marked as an undone edit. If the algorithm found such a pattern, it would mark the currently analyzed version as a revert of undo and introduce no other changes. This pattern of edits can be seen in the sequence of versions 271–274 in **Table 2.7**. Version 271 was a regular edit, followed by edit 272, which was subsequently undone in version 273. Version 274, which reverted the undo in version 273, was identical to version 272. We will therefore code version 274 as a revert of undo.

Finally, the algorithm identified whether the revert of undo was undertaken by the same person who had authored the undone edit or by someone else. To do so, the algorithm compared the usernames of the editor who reverted the undo and the editor who saved the undone edit. If the names differed, as in the case of versions 272 and 274, the algorithm would tag the revert of undo as undertaken by other. If the two names were identical, as in versions 276 and 278, the algorithm would tag the revert of undo as undertaken by self.

APPENDIX B. INTERVIEW SCHEDULE.

Starting point

- How long were you a contributor to Wikipedia?

- When did you begin contributing?

- What form did your first contributions take (edits, new articles, etc)? What was the first article you edited about?

- Why did you start editing? What was your primary motivation?

Contribution patterns

- How did your contribution pattern evolved over time? Did you edit more later or when you first started?

- How much time did you put into editing Wikipedia articles?

- Did Wikipedia play a part in your daily life? How and for how long?

Likes / dislike / departure from Wikipedia

- What, if anything, do you appreciate / like regarding Wikipedia?

- What, if anything, would you most like to change about Wikipedia?

- Why did you withdraw from Wikipedia? (if applicable)

- What, if anything, could make you return? (if applicable)

Community

- Did you feel a sense of community within Wikipedia? If so, please elaborate.

- Did you talk with other users on Wikipedia?

- Do you or did you talk to your 'real life' friends about Wikipedia? Do they contribute as well?

Offline life

- Finally, if you wouldn't mind telling me a little bit about yourself… What is your educational background?

- What is your employment status?

- What is your marital status?

- What are your primary hobbies and interests?

- I'd also like to hear a bit about your social identity. What are your friends like?

- What do you like to do with them?

- Are they similar or dissimilar to you?

- Are they a close-knit circle or a collection of dispersed individuals?

REFERENCES

Ahuja, Gautam. 2000. "Collaboration Networks, Structural Holes and Innovation: A Longitudinal Study." *Administrative Science Quarterly* 45:425-455.

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

Allison, Paul and Richard 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.

Anthony, Denise L., Sean W Smith, and Timothy Williamson. 2009. "Reputation and Reliability in Collective Goods: The Case of the Online Encyclopedia Wikipedia." *Rationality and Society* 21:283-306.

Arrow, Kenneth J. 1977. "Toward a Theory of Price Adjustment." Pp. 380-390 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.

Atkinson, Robert. 1998. *The Life Story Interview*. London, UK: Sage Publications.

Axelrod, Robert. 1986. "An Evolutionary Approach to Norms." *American Political Science Review* 80(4):1095-1111.

Bacharach, Michael and Diego Gambetta. 2001. "Trust in Signs." Pp. 148-184 in *Trust in Society*, edited by Karen 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.

Barker, James. 1993. "Tightening the Iron Cage: Concertive Control in Self-Managing Team." *Administrative Science Quarterly* 38:408-37.

Baron, Robert, Norbert Kerr, and Norman Miller. 1992. *Group Process, Group Decision, Group Action*. Pacific Grove, CA: Brooks/Cole Publishing Company.

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

Bendor, Jonathan and Piotr Swistak. 2001. "The Evolution of Norms." *American Journal of Sociology* 106:1493-1545.

Benkler, Yochai. 2006. *The Wealth of Networks : How Social Production Transforms Markets and Freedom*. New Haven, CT: Yale University Press.

Bielby, William and Denise Bielby. 1989. "Family Ties: Balancing Commitments to Work and Family in Dual Earner Households." *American Sociological Review* 54:776-789.

Biggart, Nicole. 2001. "Banking on Each Other: the Situational Logic of Rotating Savings and Credit Associations." *Advances in Qualitative Organization Research* 3:129-153.

Blake, Judith and Kingsley Davis. 1964. "Norms, Values, and Sanctions." in *Handbook of Modern Sociology*, edited by R. Faris. Chicago, IL: Rand McNally.

Bolino, Mark C., William H. Turnley and James M. Bloodgood. 2002. "Citizenship Behavior and the Creation of Social Capital in Organizations." *Academy of Management Review* 27:505–522.

Borgatti, Stephen and Pacey Foster. 2003. "The Network Paradigm in Organizational Research: A Review and Typology." *Journal of Management* 29:991-1013.

Brass, Daniel J., Kenneth D. Butterfield, and Bruce C. Skaggs. 1998. "Relationships and Unethical Behavior: A Social Network Perspective." *The Academy of Management Review* 23(1):14-31.

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.

Buriol, Luciana, Carlos Castillo, Debora Donato, Stefano Leonardi, and Stefano Millozzi. 2006. "Temporal Evolution of the Wikigraph." Pp. 45-51 in *Web Intelligence Conference*: IEEE CS Press.

Burt, Ronald S. 1982. *Toward a Structural Theory of Action: Network Models of Social Structure, Perception, and Action*. New York: Academic Press.

------. 1983. "Distinguishing Relational Contents." Pp. 35-74 in *Applied Network Analysis*, edited by Ronald S. Burt and M.J. Minor. Beverly Hills, CA: Sage.

------. 1992. *Structural Holes: The Social Structure of Competition.* Cambridge, MA: Harvard University Press.

------. 2000. "The Network Structure of Social Capital." In *Research in Organizational Behavior*, edited by Robert I. Sutton and Barry M. Staw. Greenwich, CT: JAI Press.

------. 2001. "Attachment, Decay, and Social Network." *Journal of Organizational Behavior* 22:619–643.

------. 2005. *Brokerage and Closure: An Introduction to Social Capital*: Oxford University Press.

------. 2011. "Structural Holes in Virtual Worlds." Working Paper retrieved in Febuary 2012 from http://faculty.chicagobooth.edu/ronald.burt/research/files/SHVW.pdf

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.

Carr, Deborah. 2007. "The Global Digital Divide." *Contexts* 6(3): 58.

Chwe, Michael Suk-Young. 1999. "Structure and Strategy in Collective Action." *American Journal of Sociology* 105:128-156.

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. 1988a. "Social Capital in the Creation of Human Capital." *American Journal of Sociology* 94:S95-S120.

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

------. 1990. *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.

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

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

Deci, Edward L. and Richard M. Ryan. 2000. "The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior." *Psychological Inquiry* 11:227-268.

Durkheim, Emile. 1984 [1893]. *The Division of Labor in Society*. New York: Free Press.

------. 1951. *Suicide: A Study in Sociology*. Translated by John A. Spaulding and George Simpson. New York: The Free Press.

Eby, Lillian T., Wendy J. Casper, Angie Lockwood, Chris Bordeaux and Andi Brinley. 2005. "Work and Family Research in IO/OB: Content Analysis and Review of the Literature (1980-2002)." *Journal of Vocational Behavior* 66: 124-197.

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

Eisenberg, Eric M., Peter R. Monge and Katherine I. Miller. 1983. "Involvement in Communication Networks as a Predictor of Organizational Commitment." *Human Communication Research* 10:179-201.

Ellickson, Robert C. 1991. *Order without Law: How Neighbors Settle Disputes*. Cambridge, MA: Harvard University Press.

------. 2001. "The Evolution of Social Norms: A Perspective from the Legal Academy." Pp. 35-75 in *Social Norms*, edited by M. Hechter and K.-D. Opp. New York, NY: Russell Sage.

Elster, Jon. 1989. *The Cement of Society: A Study of Social Order*. New York: Cambridge University Press.

------. 2003. "Coleman on Social Norms." *Revue Française de Sociologie* 44:297-304.

Emerson, Richard M. 1962. "Power-Dependence Relations." *American Sociological Review* 27:31-41.

Emirbayer, Mustafa and Jeff Goodwin. 1994. "Network Analysis, Culture, and the Problem of Agency." *American Journal of Sociology* 99:1411-1454.

Fehr, Ernst and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 415:137-140.

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.

Flache, Andreas and Michael W. Macy. 1996. "The Weakness of Strong Ties: Collective Action Failure in a Highly Cohesive Group." *Journal of Mathematical Sociology* 21:3-28.

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.

Freeman, Stephanie. 2007. "The Material and Social Dynamics of Motivation: Contributions to Open Source Language Technology Development." *Science Studies* 20(2):55-77.

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

Glaser, Barney G. and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Piscataway, NJ: Aldine Transaction.

Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. New York: Anchor Books.

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

Gould, Roger V. 1993. "Collective Action and Network Structure." *American Sociological Review* 58:182-196.

Granovetter, Mark. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78:1360-1380.

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

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

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.

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.

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

Greenhaus, Jeffrey H. and Nicholas J. Beutell. 1985. "Sources of Conflict between Work and Family Roles." *Academy of Management Review* 10:76-88.

Greif, Avner. 1989. "Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders." *Journal of Economic History* 49:857-882.

Gronbjerg, Kirsten A. . 1989. "Communities and Nonprofit Organizations: Interlocking Ecological Systems." Pp. 35-69 in *Dimensions of Community: A Research Handbook* edited by D. A. Chekki. New York: Garland Publishing.

Guimaraes, Paulo. 2008. "The Fixed Effects Negative Binomial Model Revisited." *Economics Letters* 99:63-66.

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

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

Hackman, Richard J. 2003. "Learning More from Crossing Levels: Evidence from Airplanes, Orchestras, and Hospitals." *Journal of Organizational Behavior* 24:1-18.

Hardin, Russell. 1971. "Collective Action as an Agreeable n-Prisoners' Dilemma." *Behavioral Science* 16:472-481.

------. 1982. *Collective Action*. Baltimore, MD: Johns Hopkins University Press.

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

Hausman, Jerry A., Bronwyn H. Hall, and Zvi Griliches. 1984. "Econometric Models for Count Data with an Application to the Patents–R & D Relationship." *Econometrica* 52:909-938.

Hayek, F.A. 2008. *Prices and Production and Other Works: F.A. Hayek on Money, the Business Cycle, and the Gold Standard*. Auburn, AL: Ludwig von Mises Institute.

Hechter, Michael and Satoshi Kanazawa. 1993. "Group Solidarity and Social Order in Japan." *Journal of Theoretical Politics* 5:455-493.

Hechter, Michael. 1987. *Principles of Group Solidarity*. Berkeley, CA: University of California Press.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47:153–161.

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.

Homans, George C. 1950. *The Human Group*. New York, NY: Harcourt, Brace.

Horne, Christine. 2001. "The Enforcement of Norms: Group Cohesion and Meta-Norms." *Social Psychology Quarterly* 64:253-266.

------. 2004. "Collective Benefits, Exchange Interests, and Norm Enforcement." *Social Forces* 82:1037-1062.

------. 2007. "Explaining Norm Enforcement." *Rationality and Society* 19:139-170.

Ibarra, Herminia, Martin Kilduff, and Wenpin Tsai. 2005. "Zooming In and Out: Connecting Individuals and Collectivities at the Frontiers of Organizational Network Research." *Organization Science* 16:359-371.

Ilies, Remus, Kelly M Schwind, David T. Wagner, Michael D. Johnson, D. Scott DeRue and Daniel R. Ilgen. 2007. "When Can Employees Have a Family Life? The Effects of Daily Workload and Affect on Work-Family Conflict and Social Behaviors at Home." *Journal of Applied Psychology* 92(5):1368-1379.

Iverson, Roderick D. 1999. "An Event History Analysis of Employee Turnover: The Case of Hospital Employees in Australia." *Human Resource Management Review* 9:397–418.

Iverson, Roderick D. and Jacqueline A. Pullman. 2000. "Determinants of Voluntary Turnover and Layoffs in an Environment of Repeated Downsizing Following a Merger: An Event History Analysis." *Journal of Management* 26(5):977-1003.

Kahn, William A. 1998. "Relational Systems at Work." In *Research in Organizational Behavior*, edited by L. L. Cummings and Barry M. Staw 20:39 –76. Greenwich, CT: JAI Press.

Kittur, Aniket, and Robert E. Kraut. 2008. "Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination." *Computer-Supported Cooperative Work Proceedings*.

Kittur, Aniket, Bongwon Suh, Bryan A. Pendleton, and Ed H. Chi. 2007. "He Says, She Says: Conflict and Coordination in Wikipedia." in *Computer Human Interaction 2007*, Proceedings of the SIGCHI conference on Human Factors in Computing Systems. San Jose, CA.

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.

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.

Knutson, Brian. 2004. "Sweet Revenge?" *Science* 305:1246-1247.

Kollock, Peter. 1998. "Social Dilemmas: The Anatomy of Cooperation." *Annual Review of Sociology* 24:183-214.

Kopelman, Richard E., Jeffrey H. Greenhaus, and Thomas F. Connolly. 1983. "A Model of Work, Family, and Interrole Conflict: A Construct Validation Study." *Organizational Behavior and Human Performance* 32:198-215.

Kossinets, Gueorgi and Duncan J. Watts. 2009. "Origins of Homophily in an Evolving Social Network." *American Journal of Sociology* 115:405-450.

Krackhardt, David and Lyman W. Porter. 1985. "When Friends Leave: A Structural Analysis of the Relationship between Turnover and Stayers' Attitudees." *Adminstrative Science Quarterly* 30(2):242-261.

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.

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.

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.

Lave, Jean and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation (Learning in Doing: Social, Cognitive & Computational Perspectives)*. New York: Cambridge University Press.

Leiter, Michael and Christina Maslach. 1988. "The Impact of Interpersonal Environment on Burnout and Organizational Commitment." *Journal of Organizational Behavior* 9(4):297-308.

Lin, Nan. 2001. *Social Capital: A Theory of Social Structure and Action* Cambridge University Press.

Lounsbury, John W. and Linda L. Hoopes. 1986. "A Vacation from Work: Changes in Work and Nonwork Outcomes." *Journal of Applied Psychology* 71(3): 395-401.

Macy, Michael W. 1991. "Chains of Cooperation: Threshold Effects in Collective Action." *American Sociological Review* 56(6):730-747.

Marwell, Gerald and Pamela Oliver. 1984. "Collective Action Theory and Social Movements Research." *Research in Social Movements, Conflict and Change* 7:1-27.

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

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

McGehee, Charles L. 1982. "Spiritual Values and Sociology: When We Have Debunked Everything, What Then." *The American Sociologist* 17(1):40-46.

McPherson, J. Miller, Pamela A. Popielarz and Sonja Drobnic. 1992. "Social Networks and Organizational Dynamics." *American Sociological Review* 57:153–170.

Menchik, Daniel A. and Xiaoli Tian. 2008. "Putting Social Context into Text: The Semiotics of E-Mail Interaction." *American Journal of Sociology* 114:332-70.

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

Mesmer-Magnus, Jessica R. and Vish C. Viswesvaran. 2005. "Convergence between Measures of Work-to-Family and Family-to-Work Conflict: A Meta-Analytic Examination. *Journal of Vocational Behavior* 67:215-232.

Molm, Linda D. 1997. *Coercive Power in Social Exchange*. Cambridge, UK: University of Cambridge Press.

Morgan, Stephen L. and Aage B. Sørensen. 1999. "Parental Networks, Social Closure, and Mathematics Learning: A Test of Coleman's Social Capital Explanation of School Effects." *American Sociological Review* 64:661-681.

Morris, James H. and J. Daniel Sherman. 1981. "Generalizability of an Organizational Commitment Model." *Academy of Management Journal* 24(3):512-526.

Mossholder, Kevin W., Randall P. Settoon and Stephanie C. Henagan. 2005. "A Relational Perspective on Turnover: Examining Structural, Attitudinal and Behavioral Predictors." *Academy of Management Journal* 48(4):607-618.

Norris, Pippa. 2001. *Digital Divide: Civic Engagement, Information Poverty and the Internet World-Wide.* Cambridge, MA: Cambridge University Press.

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

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.

Oliver, Pamela. 1980. "Rewards and Punishments as Selective Incentives for Collective Action: Theoretical Investigations." *American Journal of Sociology* 85:1356-1375.

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

------. 1993. "Formal Models of Collective Action." *Annual Review of Sociology* 19:271-300.

Olson, Mancur. 1971. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press.

Opp, Karl-Dieter. 1982. "The Evolutionary Emergence of Norms." *British Journal of Social Psychology* 21(2): 139-149.

Ortega, Felipe. 2009. "Wikipedia: A Quantitative Analysis." Unpublished doctoral dissertation at 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*.

Parsons, Talcott. 1953. "A Revised Analytical Approach to the Theory of Social Stratification." Pp. 99-128 in *Class, Status and Power*, edited by R. Bendix and S. Lipset. New York, NY: Free Press.

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

Pfaff, Steven and Hyojoung Kim. 2003. "Exit-Voice Dynamics in Collective Action: An Analysis of Emigration and Protest in the East German Revolution." *American Journal of Sociology* 109:401-444.

Piskorski, Mikolaj Jan. 2010. "Networks as Covers: Evidence from an Online Social Network." *Harvard Business School Working Papers*.

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

Portes, Alejandro and Julia Sensenbrenner. 1993. "Embeddedness and Immigration: Notes on the Social Determinants of Economic Action." *American Journal of Sociology* 98(6):1320-1350.

Randall, Donna M. 1988. "Multiple Roles and Organizational Commitment." *Journal of Organizational Behavior* 9(4):309-317.

Raub, Werner and Jeroen Weesie. 1990. "Reputation and Efficiency in Social Interactions: An Example of Network Effects." *American Journal of Sociology* 96:626-654.

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

Rhoades, Linda and Robert Eisenberger. 2002. "Perceived Organizational Support: A Review of the Literature." *Journal of Applied Psychology* 87:698 –714.

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.

Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." *Science* 277:918-924.

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

Schotter, Andrew. 1982. *The Economic Theory of Institutions*. Cambridge, UK: Cambridge University Press.

Schradie, Jen. 2011. "The Digital Production Gap: The Digital Divide and Web 2.0 Collide." *Poetics* 39(2):145-168.

Shah, Sonali. 2000. "Sources and Patterns of Innovation in a Consumer Products Field: Innovations in Sporting Equipment." *MIT Sloan Working Paper #4105*.

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

Simmel, Georg. 1902. "The Number of Members as Determining the Sociological Form of the Group.I." *American Journal of Sociology* 8:1-46.

-------. 1902. "The Number of Members as Determining the Sociological Form of the Group. II." *American Journal of Sociology* 8(2):158-196.

-------. 1964. *Conflict & The Web of Group-Affiliations*. Translated by Kurt H. Wolff and Reinhard Bendix. New York: The Free Press.

Simpson, Bill and Michael W. Macy. 2004. "Power, Identity, and Collective Action in Social Exchange." *Social Forces* 82:1373-1409.

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.

Sparrowe, Raymond T., Robert C. Liden, Sandy J. Wayne and Maria L. Kraimer. 2001. "Social Networks and the Performance of Individuals and Groups." *Academy of Management Journal* 44:316–325.

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.

Staw, Barry M. and Jerry Ross. 1989. "Understanding Behavior in Escalation Situations." *Science* 246(4927):216-220.

Strauss, Anselm and Juliet Corbin. 1998. *Basics of Qualitative Research*. Thousand Oaks, CA: Sage.

Strauss, Anselm. 1987. *Qualitative Research for Social Scientists*. Cambridge, UK: Cambridge University Press.

Sugden, Robert. 1986. *The Economics of Rights, Co-Operation and Welfare*. Oxford: Basil Blackwell.

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.

Uehara, Edwina. 1990. "Dual Exchange Theory, Social Networks, and Informal Social Support." *American Journal of Sociology* 96:521-557.

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.

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

Van de Ven, Andrew. 1992. "Suggestions for Studying Strategy Process: A Research Note." *Strategic Management Journal* 13:189-191.

Verbrugge, Lois M. 1979. "Multiplexity in Adult Friendships." *Social Forces* 57(4):1286-1309.

Weber, Max. 1976. *The Protestant Ethic and the Spirit of Capitalism*. Translated by T. Parsons and R. H. Tawney. New York: Scribner.

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.

Weiss, Robert. 1994. *Learning from Strangers*. New York: The Free Press.

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

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

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.

Willer, Rob. 2009. "Groups Reward Individual Sacrifice: The Status Solution to the Collective Action Problem." *American Sociological Review* 74:23–43.

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

Wilson, John and Marc Musick. 1997. "Who Cares? Toward an Integrated Theory of Volunteer Work." *American Sociological Review* 62:694-713.

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*.

Yamagishi, Toshio and Karen S. Cook. 1993. "Generalized Exchange and Social Dilemmas." *Social Pscyhology Quarterly* 56:235-248.

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.