MIS Quarterly

RESEARCH ARTICLE

An Attraction–Selection–Attrition Theory of Online Community Size and Resilience[1]

Brian S. Butler
College of Information Studies and Robert H. Smith School of Business, University of Maryland,
College Park, MD 20742 U.S.A. {bsbutler@umd.edu}

Patrick J. Bateman
Williamson College of Business Administration, Youngstown State University, One University Plaza,
Youngstown, OH 44515 U.S.A. {pjbateman@ysu.edu}

Peter H. Gray
McIntire School of Commerce, University of Virginia, Rouss & Robertson Halls, East Lawn,
Charlottesville, VA 22904 U.S.A. {peter@virginia.edu}

E. Ilana Diamant
College of Engineering, University of Washington,
Seattle, WA 98195 U.S.A. {ilanadm@uw.edu}

Online discussion communities play an important role in the development of relationships and the transfer of knowledge within and across organizations. Their underlying technologies enhance these processes by providing infrastructures through which group-based communication can occur. Community administrators often make decisions about technologies with the goal of enhancing the user experience, but the impact of such decisions on how a community develops must also be considered. To shed light on this complex and under-researched phenomenon, we offer a model of key latent constructs influenced by technology choices and possible causal paths by which they have dynamic effects on communities. Two important community characteristics that can be impacted are ***community size*** *(number of members) and* ***community resilience*** *(membership that is willing to remain involved with the community in spite of variability and change in the topics discussed). To model community development, we build on attraction–selection–attrition (ASA) theory, introducing two new concepts:* ***participation costs*** *(how much time and effort are required to engage with content provided in a community) and* ***topic consistency cues*** *(how strongly a community signals that topics that may appear in the future will be consistent with what it has hosted in the past). We use the proposed ASA theory of online communities (OCASA) to develop a simulation model of community size and resilience that affirms some conventional wisdom and also has novel and counterintuitive implications. Analysis of the model leads to testable new propositions about the causal paths by which technology choices affect the emergence of community size and community resilience, and associated implications for community sustainability.*

Keywords: Online communities, social media, benefits, costs, emergent systems, simulation

[1]Suprateek Sarker was the accepting senior editor for this paper. Sid Huff served as the associate editor.

The first three authors contributed equally to this paper.

Introduction

Online discussion communities are groups of people with shared interests who communicate over the Internet through a common platform (e.g., Butler 2001; Ridings and Gefen 2004). They first appeared with the widespread adoption of integrated computing and communication technologies (Rheingold 1993) and have become important channels for facilitating group discussions within and across organizational boundaries (Baldwin and von Hippel 2011). Through community platforms such as forums, newsgroups, bulletin boards, wikis, and blogs, online discussion communities have been used to support knowledge management initiatives (Hall and Graha 2004), promote brand loyalty (e.g., Bagozzi and Dholakia 2006; Schau et al. 2009), and enable buyers and sellers to form trading groups (Kollmann and Krell 2011). Their continued evolution in form, function, and underlying technologies offers researchers and practitioners new opportunities to explore and understand their structures, processes, and impacts (Zammuto et al. 2007).

As IS professionals have made choices about the technologies that enable online discussion communities, they have discovered that the emergent nature of such communities makes them challenging to develop, manage, and sustain over time. Two key characteristics of communities that relate directly to their sustainability are *community size* (the number of members[2]) and *community resilience* (the extent to which a community's membership is willing to remain involved in the face of variability and change in topics discussed). Community size provides an indication of resources available to members; in order to generate net benefits for members, communities must attract a critical mass of members who are willing to engage even if every single interaction does not produce value for them personally (Butler 2001; Wasko and Faraj 2005). Many communities stall because they cannot attract enough members to sustain themselves (Cummings et al. 2002), even when organizations invest significant resources in building a member base (Worthen 2008). Community resilience can also be problematic, with many members never returning after their first post (Arguello et al. 2006; Ducheneaut 2005), perhaps because they formed inaccurate judgments about the community's value to them (Jina et al. 2010). Over time, communities that lack resilience may collapse if discussion topics change significantly. How choices about the technologies through which a community operates may—intentionally or not—affect the size and resilience of its membership remains a critical and unanswered question for IS researchers and professionals.

In the remainder of this paper, we propose a theory to explain how discussion communities evolve in ways that are more or less sustainable. Following established best practices for developing theory using simulation methods (Davis et al. 2007), we build on attraction–selection–attrition (ASA) theory (Schneider et al. 1995), introducing two new concepts: *participation costs* (how much time and effort are required to engage with content provided in a community) and *topic consistency cues* (how strongly a community signals that the topics that may appear in the future will be consistent with what it has hosted in the past). Our goal is to better explain how technology choices that change users' participation experiences may also affect community sustainability. We examine the proposed theory using a simulation model that we calibrate and validate using data from 192 listserv-based communities. We use the model to conduct a series of virtual experiments and derive a number of propositions, and then discuss the implications of these propositions for the study and practice of online community design.

Theoretical Background

Whether their purpose is to share knowledge, provide relational support, facilitate professional practice, or develop a market for a product, online communities rely on members' continued involvement to generate benefits for each other and for the community as a whole (Butler 2001; Wasko and Faraj 2005). Although some communities are able to build large, engaged member bases, most do not (Preece 2001). Many struggle to attract and retain members (e.g., Cummings et al. 2002; Joyce and Kraut 2006; Ren et al. 2012), which limits their sustainability (Butler et al. 2007). Because a community's ability to generate benefits depends on maintaining a sufficiently large and engaged membership pool (Butler 2001; Wasko and Faraj 2005), its success depends on whether it can attract members who want to remain (e.g., Arguello et al. 2006; Ma and Agarwal 2007; Ransbotham and Kane 2011). As described in detail below, community size and resilience are two key indicators of community sustainability that reflect how effective a community has been at building such a member base.

Outcomes related to community size have often been cast as proxies for the success of an online community (e.g., Arguello et al. 2006; Lazar and Preece 2003; Ma and Agarwal 2007). Online communities provide very limited value if their size is

[2]We define membership as engaging with a community in such a way that an individual is exposed to the communication activity and resources generated by other members. This behavioral definition is useful despite its omission of psychological aspects of community involvement (Blanchard and Markus 2004) as many of the benefits of communities (e.g., knowledge transfer, social support, information awareness) are predicated first and foremost on the assumption that individuals, at the very least, are exposed to the community's activities.

too small (Blanchard and Markus 2004; Peddibhotla and Subramani 2007; Ridings and Gefen 2004) because members derive benefit from the contributions made by other members (Farrell and Saloner 1985; Katz and Shapiro 1985). Indeed, by some estimates, the value a community can provide its membership is exponentially related to the size of its membership (Spagnoletti and Resca 2012). Community administrators understand the importance of community size (Johnston 2007) as a proxy for the resources available to a community for accomplishing its collective goals (Butler et al. 2007). Large size can signal to both members and nonmembers that a community is active and vibrant (Markus 1987) and can implicitly affirm its value (Bateman, Gray, and Butler 2011; Ren et al. 2012).

Researchers have recommended that communities display information related to community size, such as public acknowledgment when a new member joins or showing a running tally of membership size (Kraut and Resnick 2011; Resnick et al. 2011). Consistent with the literature on incentives and group size (e.g., Zhang and Zhu 2011), members are more likely to contribute to a community when they believe they are addressing a large audience (Burke et al. 2009), although the impact of size on contribution rates may depend in part on personality characteristics (Nov and Arazy 2013). However, design interventions that positively affect message contribution may not increase community size (e.g., Choi et al. 2010). Increases in community size can actually have a negative effect on an individual's willingness to contribute (Jones et al. 2004; Kraut and Resnick 2011). For these reasons, it is important to investigate other causal connections by which technology features might impact community size (Ren et al. 2012).

If community size is a proxy for the benefits a member can obtain (Butler 2001; Butler et al. 2007), then providing members with indicators of community size is likely to facilitate membership growth and sustainability only when community size is sufficient—that is, when there are enough members to contribute valuable content (Resnick et al. 2011). In addition to providing an adequate pool of informational benefits (Spagnoletti and Resca 2012), a sufficiently large pool of members is also necessary to develop the social structures and norms that frame interactions, roles, contribution expectations, and community governance (Choua et al. 2010; Kane et al. 2009). However, community size cannot be directly controlled; rather, it is a result of many individual choices based on members' beliefs about the benefits of participating (Butler et al. 2007). Playing an important role in a community's long-term viability, community size is, therefore, an emergent characteristic that is driven by the interplay between technological features, member activity, and individual choice.

In addition to identifying the important role of community size, prior research also suggests that developing a membership base that is willing to remain engaged is key to community success (Bateman, Gray, and Butler 2011; Ren et al. 2012). However, online communities are often characterized by considerable membership instability (Ransbotham and Kane 2011), which negatively impacts a community's ability to provide benefits to its members (Jina et al. 2010). Online communities are unlike traditional organizations, where the occasional arrival or departure of members gradually changes the organization's make-up, focus, and benefits (Kuk 2006). Instead, online communities might best be seen as "fluid objects" (Faraj et al. 2011) in which tensions around the alignment (or misalignment) of members' interests drive never-ending cycles of change. Discussion topics thus evolve as new individuals join and old members leave (Kim 2000), as controversies erupt and subside, and as external events create interest in topics that later fades (Hummel and Lechner 2002).

A resilient membership takes such topic variation in stride (Kraut and Resnick 2011), with enough members remaining engaged even as discussion topics change. Among successful communities, these constant changes produce no real discontinuity (Law and Singleton 2005) as members adapt their attention and interests to that of the collective (Faraj et al. 2011). However, communities that lack a resilient membership may suffer as the benefits they can offer may be little more than the contributions of the newest members (Kane and Alavi 2007), which can dissolve into nothing more than random interactions (Faraj et al. 2011). Community mortality can increase, either gradually or suddenly (Farnham et al. 2000), when a membership is not sufficiently resilient. This often creates difficult challenges for community administrators, as community resilience is not subject to direct intervention; rather, it emerges from the membership's aggregated and interdependent communication behaviors, expectations, and choices, which are themselves situated within and influenced by the technological structures through which a community functions.

A range of literature has touched on the issue of enhancing community size and community resilience without the benefit of a strong underlying theory base. For instance, practitioner-oriented guides have provided advice about how to craft the features of a community to increase members' likelihood of joining and remaining (Kim 2000; Preece 2000). Researchers have provided rich descriptions of the characteristics of different online communities, and of member reactions to design features (Beenen et al. 2004; Lazar and Preece 2003; Phang et al. 2009; Rashid et al. 2006; Shen and Khalifa 2009). While beneficial, such sources of guidance generally lack coherent theories to explain or predict how changes to community platforms affect the development of a sufficiently large and resilient membership at a community level (Ren et al. 2012).

One stream of research has made solid theoretical contributions to the literature by narrowing complex community phenomena into problems that are tractable at an individual level of analysis. Studies have identified many reasons why individuals participate in online communities, including being attracted to community benefits (Blanchard and Markus 2004; Ridings and Gefen 2004), enjoying being perceived as a member (Lakhani and von Hippel 2003), and having a desire to help the community (Constant et al. 1996; Wasko and Faraj 2000). Participation may also be driven by deeper needs, beliefs, or dispositions, including a sense of reciprocity (Hall and Graha 2004; Wasko and Faraj 2005), altruism (Lakhani and von Hippel 2003; Wasko and Faraj 2005), empathy (Preece 1999), a desire for friendship (Ridings and Gefen 2004), or sense of connection with the community (Bateman, Gray, and Butler 2011; Ren et al. 2012). Casting participation as a decision that is driven by individual-level cognitions made by each member has led to important contributions, but primarily at an individual level of analysis. Community-level outcomes such as size and resilience have been, for the most part, theoretically neglected.

Prominent online community researchers have pointed to the need for research that models the emergent properties of online communities, reflecting how interests within a community change and how such fluctuations impact the direction, focus, and flow of members (Faraj et al. 2011). While such community dynamics have been of theoretical interest in previous research (Butler 2001; Faraj and Johnson 2011; Oh and Jeon 2007), the role of community platform characteristics has not been well theorized. Although some studies have sought to explain *why* sustainable online communities are able to survive and succeed (Butler 2001; Jones et al. 2008; Jones et al. 2004), they offer little insight into questions of *how* technology choices may affect emergent community outcomes (Benbasat and Zmud 2003). Community researchers (Zammuto et al. 2007) have called for more research to explore and theorize the "black box" of technology in a community context, recognizing the intertwining between technology, individual member decisions, and higher-order community characteristics. Others have pointed to the need for better understanding of the relationship between emergent online community characteristics and the technological platforms by which communities operate (Faraj et al. 2011). To provide a strong conceptual foundation for moving this work forward, we propose a theory and develop a model that explains how online community size and resilience can be parsimoniously explained through member attraction, selection, and attrition.

An ASA Theory of Online Communities

Online communities are in some ways quite different from traditional organizations, but both evolve as a function of the constantly changing cast of characters that are their members. Attraction–selection–attrition (ASA) theory (Schneider 1987; Schneider et al. 1995) models the emergence of organizational characteristics as a function of three processes that reflect members' decisions to join, remain in, and ultimately leave an organization. Our proposed ASA theory of online community dynamics (which we term OCASA) begins with the premise that these same three processes are central to the emergence of discussion community characteristics.

Attraction

Prospective employees are attracted to organizations based on organizational characteristics they infer but cannot fully know in advance (Turban and Keon 1993). Depending on the information they can gather, some will be more likely than others to conclude that an organization is attractive (Schneider et al. 1995). Differences in perceived attractiveness are also influenced by individuals' varied preferences regarding employment conditions, job characteristics, and organizational features (Judge and Bretz 1992; Vroom 1966). Alignment between individuals' interests and the organizational information they obtain from published sources and interpersonal contacts makes an organization more attractive to them. As a result, they are more likely to seek to join it (Rynes et al. 1991).

Analogous processes operate in the online community ecosystem. Potential members are attracted to a community by their expectations of congruency between their personal interests and the community's topics, goals, and activities (e.g., Ridings and Gefen 2004; Wasko and Faraj 2005). Much depends on what we term *initial fit expectations*, or the expected degree to which an individual believes that his or her interests match a community's discussion activity. Potential members may develop such expectations by reading formal descriptions of a community (e.g., "About Us"), official documentation (e.g., FAQs), or prior communications (e.g., archived discussion posts). When initial fit expectations are high, potential members believe that they will benefit from joining a community. Because their decision to join hinges critically on the anticipated benefits of doing so, those who do become members arrive with positive expectations that the content of community discussions will match their interests (Levine and Moreland 1999; Wanous et al. 1992), even though they may in fact have made incorrect inferences about the true fit with their interests (Joyce and Kraut 2006; Turner et al. 2005).

Selection

The selection process picks up after individuals join an organization and involves ongoing evaluation of congruence between their interests and the organizations based on actual experience (Ryan et al. 2000). Things like work assignments, enacted values, and work climate provide new information that cause members to reevaluate congruence. This continual reassessment may change how each individual engages with the organization. A good person–organization fit is associated with increased job satisfaction and commitment (Cable and Judge 1996). Conversely, perceived incongruence in interests or values can lead to dissatisfaction, reevaluation of the person–organization relationship, and, ultimately, turnover (Posner 1992).

A similar selection process takes place in online communities as members' direct exposure to discussions after joining changes their fit expectations. Consistent with work on individual–group socialization, these expectations evolve as a function of the benefits individuals believe they stand to gain by engaging a group (Moreland and Levine 1982). A range of benefits may be derived from message-based discussions, including access to information (Galegher et al. 1998), support (Ridings and Gefen 2004), and access to expert advice (Lampel and Bhalla 2007). In accordance with general theories of benefit evaluation and belief updating (e.g., Brinthaupt et al. 1991; Carley 1991; Turner 1988), members' first-hand exposure to a community's discussions leads them to update their expectations about the future availability of such benefits.

Although members' initial decisions to participate are influenced by the value of the benefits they expect to obtain, their intention to remain in the community is influenced by the fulfillment of such expectations (Jina et al. 2010). However, members' initial fit expectations may differ from these updated post-joining expectations, in part because initial expectations are necessarily based on incomplete information (Ransbotham and Kane 2011). Further, the focus of online discussions continually emerges as the membership changes (Figallo 1998; Rheingold 1993). As a result, content-related benefits that are valued by a member may fade or disappear over time. Observations of past messages, therefore, provide inherently imperfect indicators of future benefits. Members' exposure to current discussion activity leads them to update their expectations about the likelihood of obtaining benefits by remaining engaged. Because individuals in a social setting do not remember every instance of discussion activity, they form generalized expectations that become weaker or stronger as they are exposed to disconfirming or supporting evidence (Wasserman and Faust 1994). Thus, members continually update their overall evaluation of the fit between a community and their interests even though they do not remember every message they have encountered.

Attrition

The attrition process refers to an employee's decision to leave an organization because of a mismatch between their expectations and that organization's characteristics (Bretz et al. 1989; Ryan et al. 2000). Employees may leave for a wide range of reasons, but ASA considers the misalignment between what they want and what organizations provide as central (Schneider 1987; Schneider et al. 1995). Attrition means forgoing the benefits associated with membership—perhaps an appealing alternative when those benefits no longer matter to an individual. Regardless, employees' departures reduce organizational size and, therefore, reduce the resources available to perform organizational activities. Over time, attrition reshapes an organization as a whole (Schneider et al. 1995) in ways that alter the benefits it can provide to those who remain.

Individuals will leave an online community when they expect that the community's future discussions will not align with their own interests (e.g., Andrews 2002; Ransbotham and Kane 2011; Ridings and Gefen 2004). Attrition alters community size and changes community characteristics; when members leave, they cease contributing, and discussion topics shift toward the interests of the remaining members. These changes, in turn, influence other members' ongoing evaluations of a community, producing follow-on departures and even more topic changes. Together, such changes significantly alter a community's long-term trajectory by altering the benefits available to members, thereby affecting a community's viability over time (Spagnoletti and Resca 2012).

ASA theory provides a useful foundation for modeling community emergence as a function of members' participation decisions that are driven by their expectations of fit. However, it provides little help in understanding how technology choices affect community size and resilience because it is missing key theoretical constructs and processes that can be directly influenced by technology. Because the rapid evolution of community technologies limits the impact of research that theorizes specific technical features, we instead theorize constructs that can be affected by a range of technical features, but are not features themselves. A feature-based model might, for example, consider the direct impact of a particular search interface or certain content preview functionality on community size and community resilience. But, as technological advances introduce new features and make older ones obsolete, such feature-specific findings might quickly fade from relevance. Our goal is instead to produce a more flexible and robust theoretical model by following what Orlikowski and Iacono (2001) term a "proxy view" of technology, providing a foundational theoretical model that captures an essential impact of IT that could be instantiated in many different ways by different technological features. We follow this approach in order to provide a general theoretical framework upon which future theoretical and empirical studies (including feature-oriented studies) can build, and through which such studies can be integrated. In what follows, we extend ASA theory by articulating the causal connections through which a community's technological features can influence ASA processes by affecting participation costs and topic consistency cues.

Participation Costs

Members derive benefit from discussion communities by navigating, searching, and reading messages (e.g., Welser et al. 2007), but these actions take considerable time and effort. Communities with large message volumes may be active and vibrant, but the costs of attending to and processing many messages are often significant deterrents to member engagement (Jones et al. 2004). A larger message volume means that members must both invest more time to keep up with discussions and also sort through more irrelevant content to find the messages that are of interest. When members expect these *participation costs* to be high in the future, they are more likely to disengage from the community (Bateman, Gray, and Butler 2011; Kuk 2006; Millen et al. 2002).

As members become involved in a community, they observe the number and frequency of messages that are typically exchanged (Kim 2000), and form *volume expectations*—the level of discussion activity an individual believes is typical in a particular online community. Volume expectations influence both individuals' initial decisions to join a community and their continued involvement (Jones et al. 2004; Lakhani and von Hippel 2003). Together with fit expectations, volume expectations affect their perceptions of the costs and benefits of continued membership.

Community technology choices can significantly change the impact of discussion volume on perceived participation costs. Community technical features affect individuals' behaviors by providing affordances that facilitate, or hinder, the performance of activities within a community (e.g., Andrews 2002; Kim 2000). These affordances raise or lower the costs of participation—that is, they change the relationship between the volume of discussion activity within a community and the costs incurred by members. For example, some technologies make it easier to find and consume content (Arguello et al. 2006; Viegas et al. 2007), develop and maintain personal profiles (Boyd and Ellison 2007), or monitor and add to ongoing discussions (Majchrzak et al. 2013). Features such as discussion threading and filtering may drastically reduce the costs of dealing with messages of interest, thereby reducing average participation costs. Other features, such as ratings, voting, and other message-level community management mechanisms can increase the impact of discussion volume on participation costs by adding visual and procedural complexity to community content. Whether they make it easier or harder for members to handle larger message volumes, technology choices alter the impact of discussion volume expectations on future engagement with a community by changing members' assessments of participation costs.

Increased participation costs offset individuals' perceptions of the benefits to be obtained when their interests fit well with the discussion activity within a community. Importantly, although there may be variation across members in how they appropriate such technologies (e.g., Dennis et al. 2001), the presence or absence of a technical feature will have a general effect across all individuals' participation costs and their evaluation of the implications of continued membership. Individuals' assessment of net benefits (benefits less costs) therefore provides an important theoretical channel by which to conceptualize the impact of technology choices on emergent community-level outcomes such as size and resilience.

Topic Consistency Cues

An important difference between traditional organizations and online communities is the extent to which they provide consistent benefits to members over time. Traditional organizations often have considerable inertia (Hannan and Freeman 1984) due to slow-changing aspects such as formal structures, processes, routines, contracts, hierarchies of decision making, and labor arrangements (Scott and Davis 2007). Members of many traditional organizations can often safely assume that the activity they experience—and hence the benefits they receive—tomorrow will be very similar to what they experienced today, because these slow-to-change aspects are visible, interlocking, and intended to produce continuity.

Members of online communities have fewer such assurances of continuity because the activity they experience arises from the voluntary actions of individuals who have more tenuous relationships with a community (Bateman, Gray, and Butler 2011), who are inconsistent in their participation levels (Joyce and Kraut 2006), and who are often invisible even when "present" (Bateman, Pike, and Butler 2011). Individuals can quickly enter a community, contribute to a discussion (even dominate it), and leave without any indication of their departure. Discussion topics may explode onto the scene and then just as quickly disappear. Sometimes these are short-term bursts, but other times they are the beginnings of a major change in direction in the topics discussed in a given community (Farnham et al. 2000). When dealing with this unavoidable uncertainty, members must make judgments about the degree to which recent discussion activity is representative of future activity. These judgments can range from complete confidence that future topics will be similar to those currently discussed to significant skepticism that recent conversation topics will be continued into the future. Ultimately, such expectations affect decisions about continuing membership because they affect the degree to which current activity impacts individuals' developing expectations about the future fit between community discussion activity and their interests.

Many technological features available in a community can provide topic consistency cues that affect how individuals experience the community's discussion activity (Beenen et al. 2004; Ludford et al. 2004). For instance, visualization features and interface functionality can emphasize or deemphasize parts of a community's discussion history (Donath et al. 1999). Features such as member profiles and post ratings may strengthen others' beliefs that current topics are more likely to be discussed in the future. Other features, such as thread view counts and thread reply counts, can affect topic consistency cues by helping members see whether or not topics are broadly interesting to others. Communities may signal a lack of topic consistency when technical features provide little contextualization of discussion content, no cues about the growth and trajectory of a discussion topic, and little or no information about the people involved and their history with the community. In this way, community platforms may lead individuals to place less weight on recent discussion activity as an indicator of future fit. While individuals will naturally vary in their topic consistency expectations for many reasons (e.g., personality, experience using communities), community technology features provide important cues that will influence those expectations systematically for all members. Ultimately, topic consistency cues influence community resilience by affecting whether a community's membership will remain engaged in the face of changing discussion topics.

Taken together, the attraction, selection, and attrition processes framed by traditional ASA theory produce a series of feedback effects that alter the topics discussed and the potential benefits a community can offer members, which in turn affects individual-level decisions to join, stay, or leave. By extending ASA theory to account for participation costs and the somewhat ephemeral nature of online communities, we provide a foundation for examining and explaining the link between individuals' beliefs about a community, their behaviors within a community, and ultimately, as we will see below, the emergence of community characteristics such as size and resilience.

Methodology

Simulation has become an integral part of scientific investigation, fueled in part by increasingly powerful computers (Hamming 1997). Simulations are simplified computational models of some, but not all, characteristics of real-world processes, systems, or events (Law and Kelton 1991), in which theoretical logic is instantiated in executable code (Davis et al. 2007). The input parameters of a simulation model are calibrated to reflect observed conditions (Lave and March 1975) and can be manipulated to represent other possible conditions, allowing researchers to conduct virtual experiments (Carley 2001) whose results can be analyzed to better understand the implications of a theory instantiated in the model (Bratley et al. 1987; Chen and Edgington 2005). This makes systematic investigation of proposed relationships possible, yielding propositions that can be used to inform future theorizing and guide empirical work.

As a theory development method, simulation affords researchers a controlled and consistent way of examining a theory's premises and exploring its implications. Simulation can enhance theoretical precision, improve internal validity, and guide theoretical elaboration (Carley 2001). Some theoretical understanding of the focal phenomena is necessary to construct a simulation model, but its true value "depends on an incomplete theoretical understanding such that fresh theoretical insights are possible from the precision that simulation enforces and the experimentation that simulation enables" (Davis et al. 2007, p. 483). Simulation is well suited for identifying assumptions and implications of the logic at the heart of developing theories (Carroll and Harrison 1998; Kreps 1990). It is a particularly useful tool for developing theory around complex phenomena (Zott 2003) such as those involving feedback loops, nonlinear processes, or high-order indirect effects (Burton and Obel 1995; Oh and Jeon 2007). Simulation researchers can explore multiple scenarios, analyze complex systems before their creation, and conduct research that would be too costly or disruptive using other methods (Bapna et al. 2003). Classic simulation studies have propelled research in new directions in areas such as decision making (Cohen et al. 1972; March 1991), organizational evolution (Herriott et al. 1985; Morecroft 1985), and organizational learning (Lant and Mezias 1990).

This study applies simulation as a theory-building strategy. We model the ASA processes by which complex and higher-order community phenomena emerge from individual behaviors, thereby extending theories that provide only a partial explanation of discussion community dynamics. We implement a model and use it to conduct a series of virtual experiments to explore emergent consequences that are not posited in the original theory and not formally tested to date. The higher order outcomes (community size and resilience) are not direct results of specified parameters, but are instead the output of complex interactions among elements of the model. Our simulation model thus provides a bridge between process explanations of community development (e.g., Ridings and Wasko 2010) and variance models of community outcomes (e.g., Butler 2001; Jones et al. 2004).

Davis et al. (2007, p. 482) prescribe a multistage approach for developing theory through simulation. Following their approach, we (1) articulated a research question (how can technology choices influence community size and resilience?) and (2) identified a "simple theory"[3] to be the foundation for understanding the phenomena of interest (attraction–selection–attrition theory). In the following sections, we continue the process by (3) describing our simulation approach; (4) creating a computational representation based on the specifications of the model; (5) verifying the computational representation by calibrating and validating the model; and (6) experimenting to build novel theory, through the analysis of simulation results and developing propositions.

Simulation Specification

To develop the OCASA theory, we implemented an agent-based model, a type of computational model well suited for studying how actions and interactions of autonomous agents affect emergent characteristics of a larger system (Carley 2001). To examine the feedback loops and higher order effects implied in the OCASA theory, we combined the agent-based model with a discrete event simulation that models the passage of time as a series of distinct periods. This approach was selected because it allows for the modeling of emergent outcomes arising from simultaneous actions of autonomous agents without the computational resource requirements and design complexity of true parallel execution (Law and Kelton 1991).

In our model, an online community consists of two types of agents: a platform and autonomous individuals. A platform is a passive agent that enables individuals to access messages from, and disseminates messages to, other members. A community has a single platform through which members communicate. Individuals are active agents who may join, contribute messages to, receive messages from, and leave a community by interacting with its platform. Through these interactions, *community size* and *community resilience* emerge. Below, we describe how OCASA theory is instantiated in the capabilities, characteristics, and interactions of individuals and a platform.

Modeling a Platform

In the model, platforms have two capabilities (membership tracking and message management) and two characteristics (participation cost and topic consistency cues) (Table 1). Membership tracking refers to a platform's ability to passively maintain a community membership list in response to individuals' joining and leaving requests. An individual choosing to join a community informs its platform and is then added to the community, gaining access to the communication capabilities of the platform (message creation, dissemination, and access). Likewise, an individual who chooses to leave informs the platform and is removed from the community.

Message management is a platform's ability to accept messages from members and disseminate them to others in the community. Each message is modeled as having a single topic (represented by a real value between 0 and 1 inclusively), but the same topic may be the subject of multiple messages. During each time period, a platform allows, but does not require, each member to submit a single message on a topic of their choice. After all members have had an opportunity to submit a message, the platform distributes the messages to all members.

Participation costs (PC) refer to the time and effort that an individual must expend in order to derive benefit from the discussion activity within an online community. Because members are sensitive to the volume of activity within an online community (Butler 2001; Jones et al. 2004; Ridings

[3]Davis and his colleagues suggest that the parsimony afforded by simple theory provides focus and precision valuable to theorizing. A simple theory is one that involves only a few constructs, with some empirical or analytic grounding but that is limited by incomplete underlying theoretical logic. It often includes concepts from well-known theories, especially when the research focuses on their interaction.

Table 1. Platform Capabilities and Parameters		
Platform Capabilities	Member Tracking	The ability to identify community members based on two passive methods: accepting new members and removing departing members
	Message Management	The ability to collect submitted messages from and distribute messages to current community members
Platform Parameters	Participation Cost (PC)	Within the community platform, the ratio of the per-message processing cost over the average benefit provided by a message deemed interesting (Range: [0,1])
	Topic consistency cues (TCC)	The degree to which a community platform signals that current message content is indicative of future content (Range: [0,1])

and Wasko 2010), we base PC on a per-message cost that is incurred by each member as they receive and attend to each message. Specifically, we express PC as a ratio of the per-message processing cost and the average benefit provided by a message deemed interesting.[4] This allows two unobservable parameters (PC and per-message benefit) to be collapsed into one, simplifying calibration, validation, and analysis. The result is a value ranging from 0 to 1[5] that captures the degree to which processing a message is more (or less) useful than the benefit obtained from doing so. While individual differences might lead to variation in participation costs, we expect these idiosyncrasies will be randomly distributed within a population. The use of a common parameter for all individuals allows for a tighter examination of the impact of community technology attributes on community outcomes. This approach is particularly useful when, as is the case in this study, the focus is on developing theory related to emergent phenomena (Burton and Obel 1995).

A second aspect of community technology that is central to community development in the proposed OCASA theory is *topic consistency cues* (TCC). Message topics will change over time, and members must decide whether changes signal that a community will be less interesting to them in the future, or are merely momentary fluctuations that have no future implications (Farnham et al. 2000). Community technologies profoundly affect how members experience discussion activity and are likely to have systematic effects on the conclusions members form about whether current discussion activity is representative of future communication activity. To capture this characteristic of community life, we model topic consistency cues as how strongly a community's technologies signal that topics that may appear in the future will be consistent with what have appeared recently.

As a passive agent, a platform does not determine who may join or what topics may be discussed. Community characteristics such as size, discussion volume, member loss, and resilience are not directly determined by the modeled platform parameters. These emergent community characteristics are affected by the capabilities and parameters of its platform, but they are only realized when individuals interact with one another through the platform. As such, the model is not technologically deterministic, but rather conceptualizes the consequences of technology as arising from the interplay elements within a complex sociotechnical system.

Modeling Individuals

We model individuals as autonomous active agents that have interests and expectations, make choices, and create messages (Table 2). As theorized above, individuals' engagement with a community follows a common pattern. They discover and join a community based on their initial expectations about it (attraction). After joining, individuals may contribute messages and update their expectations about the community in response to the messages they observe (selection). Based on their experience, they either stay or leave (attrition). While individuals all engage a community through these general processes, they differ with respect to their initial expectations, their predilection to contribute messages, and their interests. It is through the aggregation and interplay of these individual differences that community characteristics emerge.

Attraction to an online community involves three activities: discovery, initial expectation formation, and joining. There is little research that considers how individuals discover and form initial expectations regarding communities (e.g., Bateman, Gray, and Butler 2010), but this is less important to OCASA theory than what those expectations are. Therefore, we treat discovery and initial expectation formation as exogenous to the simulation.

[4]Value is arbitrarily set to 1 to facilitate interpretation of the model and results.

[5]PC values can mathematically exceed 1, but in such a situation individuals expect the cost of processing interesting messages will exceed their benefit. Under these conditions, no one will join the community. Hence PC > 1 necessarily result in a failed community.

Consistent with the theory as outlined earlier, we conceptualize individuals' initial expectations as consisting of *initial fit expectations* (F_0) and *initial volume expectations* (V_0). We represent initial fit expectations with values between 0 and 1 that indicate the degree to which an individual expects a community's discussion will match his/her interests. Low values indicate that most of a community's messages are expected to be outside an individual's interests, while high values imply the expectation that most message activity will be of interest, and hence beneficial. We represent initial volume expectations by positive values that indicate how many messages an individual expects a community will generate in a given time period. Low values indicate that an individual does not expect there to be much discussion activity, while higher values indicate an expectation of high message volumes. Together, initial fit expectations and initial volume expectations capture individual variation arising from the attraction process.

OCASA theory posits that expectations about interest fit and volume of activity will affect individuals' engagement decisions because they will be attracted to a community when they expect the benefits of involvement to outweigh the costs. Messages that are of interest provide benefit, while those outside an individual's interests do not; therefore, the expected benefit is based on the number of messages that an individual expects to be of interest (F_0*V_0). However, an individual must deal with all community messages in some way, even if they are not of interest. Thus, the expected costs of involvement are a function of both the total expected message volume (V_0) and the per-message participation cost arising from the characteristics of the platform (Table 1). Individuals evaluate a community by comparing the benefits they expect to derive from messages (F_0*V_0) with the expected cost of dealing with those messages ($PC*V_0$). Those whose initial expected net benefit ($E_0 = F_0*V_0 - PC*V_0$) is positive will be attracted to the community and will join it by submitting a request to the community platform.

Once individuals join a community, its platform provides them with the ability to send messages and derive benefit from messages sent by others. Central to these activities are individuals' interests, which constrain the topics of messages they might contribute and determine whether a received message is beneficial. Ultimately, it is the presence or absence of these benefits that leads individuals to determine their degree of fit with the community. Individuals' interests (I) are represented as a range of values between 0 and 1[6] on a unit circle.[7] An individual with broad interests might be represented as interested in topics 0.1 through 0.9 (I = [0.1, 0.9]) while one with more focused interests might be interested in topics 0.2 through 0.35 (I = [0.2, 0.35]). Individuals' interests factor into their assessment of the expected benefit from discussion activity through their fit expectations. They also provide a basis for modeling individuals' message contribution behavior by determining the range of topics about which they are capable of creating messages.

As with the development of initial expectations, the decision to contribute messages is affected by many idiosyncratic individual and contextual factors (e.g., Bateman, Gray, and Butler 2011; Ridings and Gefen 2004; Wasko and Faraj 2005). This decision is modeled as a probabilistic process based on an individual's *message contribution probability* (MCP). During each time period, each individual has an opportunity to create and disseminate a message to the community. MCP is a value between 0 and 1 that represents the likelihood an individual will contribute by randomly selecting a topic[8] from their interests, creating a message, and submitting it to the community platform. Low values indicate that an individual is unlikely to communicate; at the extreme, a MCP of 0 indicates a "lurker" who is able to contribute messages, but will never do so. A high MCP value indicates that an individual is very active, regularly contributing messages. Together, an individual's message contribution probability and interest distribution probabilistically characterize their message contribution behavior.

OCASA theory posits that individuals' evolving *fit expectations* (F_t) and message *volume expectations* (V_t) are central to their ongoing involvement in a community. Each individual's fit expectation (F_t) is a value between 0 and 1 representing their expectations at time period t about the degree to which future messages will fit his or her interests. An individual's volume expectation (V_t) is a positive value representing their expectation at time t of the number of future community discussion messages that will appear. Modeling expectations in this way is consistent with research that suggests that individuals maintain generalized memories of prior social activity, not detailed histories of past social activity (Wasserman and Faust 1994, p. 57).

[6]The representation assumes that individual interests are contiguous within the topic set (i.e., an individual is interested in all topics in their interest range and no topics outside it).

[7]Instead of being represented as a line with 0 and 1 at the extremes, the set of topics possible in a community is represented as a circle with circumference of 1. This captures the logic of a linear topic space, while avoiding edge effects associated with end points in a linear model.

[8]New message topics are selected by an individual from a uniform random distribution bounded by their interest range so all topics within the interest range are equally likely to be chosen.

Table 2. Individual Capabilities, Parameters, and Characteristics		
	Joining	Deciding to become involved in a community based on initial expectations of positive net benefits
	Message Contribution	Creating and submitting messages for distribution to a community
	Expectation Updating	Updating net benefit expectations for continued involvement based on the message activity experienced in a community
	Leaving	Ceasing involvement in a community based on the assessment that the expected net benefits are negative
Individual Parameters	Initial Fit expectations (F_0)	Expectations formed prior to joining about the degree to which an individual's interests match the discussion within a community (Range: 0 to 1)
	Initial Volume Expectations (V_0)	Expectations formed prior to joining about the number of messages in a community each period (Range: $V_0 \geq 0$)
	Message contribution probability (MCP)	The likelihood of the individual contributing a message to a community (Range: 0 to 1)
	Interests	The set of topics in which an individual is interested (a range of values between 0 and 1)
Individual Characteristics	Fit expectations (F_t)	Expectations about the future match between an individual's interests and a community's discussion activity, arising from initial fit expectations and exposure to messages (Range: 0 to 1)
	Volume Expectations (V_t)	Expectations about the number of messages that will be present in a community each period, arising from initial volume expectations and exposure to community messages
	Net Benefits Expectations (E_t)	Expectations about the net benefits of remaining involved in a community

Individuals' content and volume expectations change as they experience a community's discussion. Each message received by an individual is a resource that, if it is of interest, is a basis for deriving benefit from involvement in a community. Each message is also a signal providing information about content and amount of activity that might be expected in the future, which leads individuals to change their expectations about future message content and volume. We model fit expectation updating as a reinforcement process (Hunter et al. 1984); each observed message that is of interest leads an individual to increase expectations about the proportion of future messages that will be of interest, while each message that is not of interest leads to a corresponding decrease in expectations. The amount of that increase is determined by a platform's topic consistency cues (TCC) and a function of the individual's prior fit expectations[9] ($F_{t-1} - F^2_{t-1}$). Each message an individual encounters that is not of interest leads to a downward adjustment of future fit expectations by the same amount ($DF_t = TCC*(F_{t-1} - F^2_{t-1})$). After all messages for a given period are distributed, individuals update their volume expectations (V_t) by maintaining an running mean message volume for all periods they have been involved in a community. Together these processes operationalize the selection process described in the OCASA model by specifying particular mechanisms by which individuals adjust their expectations about a community's discussion content and volume.

Expectations are also central to the OCASA model's attrition processes, whereby individuals reevaluate their involvement and may leave a community. When individuals' expectations about the degree of fit between their interests and future community messages (F_t) change, so do their expectations about the potential benefit they may obtain by continuing to participate ($F_t * V_t$). Similarly, as expectations about the level of message activity (V_t) change, so do expectations about the associated benefits ($F_t * V_t$) and costs ($PC*V_t$). Thus, as an individual's content and volume expectations fluctuate, their expectations about the net benefit of continued involvement change as well ($E_t = F_t * V_t - PC*V_t$). If an individual's assessment of expected net benefits drops below zero, the individual submits a request to the platform to leave the community, and attrition occurs. At this point, an individual can no longer create or receive community messages and therefore no longer receives benefits or incurs costs via the community.

[9]Fit expectation updating is modeled with this functional form based on evidence that changes in individuals' beliefs and expectations are nonlinearly related to the strength of current beliefs (Hunter et al. 1984). Strong negative or positive current expectations are associated with small changes, while weaker, more ambivalent expectations are associated with larger changes in response to additional information.

Table 3. Community Characteristics	
$Size_t$	The number of members in a community at time t
Community $Resilience_t$	The degree to which members will tolerate nonbeneficial content without leaving a community. This is assessed based on average fit expectation among the members at time t.
Member $Loss_t$	The number of members choosing to leave a community at time t.
Discussion $Volume_t$	The number of messages submitted by and distributed to community members in time t.

To model a community, it is necessary to model a community platform and a population of individuals. A community platform is instantiated by specifying participation costs (PC) and topic consistency cues (TCC) values. A population of potential participants is modeled by specifying an attraction rate (A) and distribution parameters for initial fit expectations (F_0), initial volume expectations (V_0), message contribution probabilities (MCP), and interests (I) among the individuals. The attraction rate is the number of individuals that choose to join in each time period. The distributions of initial content and volume expectations describe how individuals vary with respect to the initial expectations they form about a community. The distribution of message contribution probabilities specifies how individuals differ in their tendency to contribute messages. The distribution of interests describes how individuals vary in terms of the subset of the topic space they find to be beneficial. Taken together, these parameters (see Appendix A) describe the information needed to represent an online community within the proposed model.

Characteristics of a discussion community emerge from the interplay of the community platform and the actions of individuals. Community size is the number of individuals who have chosen to remain in a community in a particular time period. Community resilience reflects individuals' overall willingness to remain in spite of variability and change in the topics discussed. This community characteristic is reflected in the mean level of fit expectations among members, which is their average expectation about the degree of fit between their interests and future content. This is an indicator of the magnitude of expectation change that would be necessary before a significant proportion of a community's members would end their involvement. Member loss is the number of members who choose to leave a community at the end of a given period. Discussion volume is the number of messages members submit to the platform in a given period. Community size, community resilience, member loss, and discussion volume are all characteristics of a specific community (Table 3) that do not directly arise from model specification, but rather from the interaction of community platform features, individual tendencies, and context dependent choices.

The multi-agent model described here was implemented using MATLAB, a high-level computer environment for computation and visualization that integrates numerical analysis and graphics. MATLAB is widely used for simulation implementation and analysis. We selected it for this project because it provides a strong combination of flexibility, power, ease of use, high quality documentation, and peer support (Attaway 2012; Mikolai and Madey 2009). This combination of features allowed for efficient implementation of the discrete event, multi-agent model specified above, calibration and validation of the model, and execution of the virtual experiments and analysis used to examine the implications of the model. The MATLAB scripts that implement the model described here are available from the authors on request.

Model Calibration and Validation

Before examining a simulation model's implications, connections must be made between the model and observed features and characteristics of the phenomena of interest (Bratley et al. 1987; Chen and Edgington 2005). Calibration grounds a model in data drawn from real-world phenomena in order to anchor model parameters, while validation tests the model to verify correspondence between modeled and observed emergent characteristics. In the calibration process, simulation parameters are set to reflect the features of empirical data (Carley 1996), resulting in parameter values that represent a particular sample of discussion communities. These parameter values provide a reference point for model validation, analysis, and interpretation. Validation of the model assesses how well the emergent characteristics of modeled samples of communities matched those of the validation subsample. These processes establish that the model as constructed mirrors critical aspects of online discussion community dynamics and hence is a reasonable basis for developing specific propositions regarding the role of technological features in their development and evolution.

For this study, calibration and validation began with a random sample of unmanaged listserv communities that communicated via mailing lists and list management software. From over 70,000 listservs that had been in operation for at least four months, we selected a random sample of 284, stratified by subject to ensure that it spanned a range of topics and member populations. For each community, all discussion messages and daily membership lists were collected for 130 days. This observation period was chosen because it was long enough to observe meaningful changes, but still feasible given constraints on the available processing and data storage resources. Communities that ceased operation, changed message formats, or altered the accessibility of member data during the data collection period were dropped, resulting in a dataset of 192 communities. From this set, 50 percent (96) were randomly selected for the calibration dataset; the remaining became the validation dataset.

Model Calibration

The model was calibrated by (1) calculating the distributions of observable parameter values among the communities within the calibration sample, (2) assigning distributions to some secondary unobservable parameters based on prior literature, and (3) estimating primary unobservable parameters based on fit between the distributions of secondary community characteristics in modeled samples of communities and those observed in the calibration sample.

The first step involved determining the distribution of observable parameters (attraction rate and message contribution probability distributions) within the calibration sample. We based the attraction rate for each community on the average number of new members per day over the entire observation period. In each step of calibration, we used probability plots (P-P) to determine whether a gamma, Weibull, exponential, or log-normal distribution was the best fit for the empirical data (Law and Kelton 1991, p. 374). We calculated the per-month rate of entrants for each community in the calibration sample, and then fit a distribution to these data, which was well described by a gamma distribution with a = 0.2624 and b = 1.1264 (Appendix A). In our model, individuals within a population discover a community probabilistically, based on its attraction rate (A) that is drawn from this distribution.

The message contribution probability (MCP) distribution for each community was also determined using raw data from the calibration sample. However, because MCP distributions are typically highly skewed (Butler 2001; Jones et al. 2004), it is not sufficient to represent them with a single measure of central tendency. Instead, the MCP distribution for each community was represented by two values: a participation ratio and a participation probability. Participation ratio is the proportion of individuals who contribute a message at least once while members (i.e., the non-lurkers). It was calculated for each community as the total number of unique contributors divided by the number of members at any time during the observation period. For the calibration sample as a whole, the distribution of participation ratios was determined to be well described by a gamma distribution with a = 0.3024 and b = 0.5268. Participation probability is the likelihood that a non-lurker will choose to contribute a message in a given period, calculated for each community as the number of messages per day per unique contributor over the observation period. In the calibration sample, participation probabilities were log-normally distributed with $\mu = -4.10$ and $\sigma = 0.55$.

The second step of model calibration involves assigning distributions for the secondary unobservable parameters (a community's interest distribution, initial fit expectations, and initial volume expectations). While it is ideal to calibrate parameters using empirical data, for some parameters this simply may not be possible. In these cases, parameter values may be set based on prior research or be inferred by matching simulation outcomes with other observable data (Law and Kelton 1991). In a process analogous to randomization in experimental design, unobservable parameters that are not the primary focus of investigation may be anchored using general distributions. This has the effect of allowing for variation in these secondary parameters while controlling for their effects on the processes and relationship of interest.

A community's interest distribution describes how individuals' interests vary. Generally, individuals' interest will differ with respect to their location in the topic space (interest location) and the amount of the topic space they cover (interest breadth). Because it is not feasible to collect such data about individuals' interests, a distribution of interest distributions was constructed so that the modeled communities would represent a range of populations. Interest locations were assumed to be uniformly random within the topic space, capturing the idiosyncrasies of individual interests in a way that is independent of any particular community. Individuals' interest breadth values were uniformly distributed between 0 and a community-specific maximum selected from a uniform random distribution bounded by 0.25 and 0.75. This resulted in simulated samples of communities that varied with respect to both the structure and the particulars of their target population's interests.

The distribution of initial fit expectations (F_0) among members joining each community was modeled as a uniform distribution between 0.75 and 1.0. This range was selected to reflect the tendency of individuals to have optimistic expectations about the value of involvement in groups that they join (Brinthaupt et al. 1991). This distribution was used for all communities because there was no *a priori* reason to expect that communities would vary with respect to the first impressions that joining individuals form. Initial volume expectations (V_0) were set to a fixed value of 1 for all communities, corresponding to the expectation of 1 message per period. This value was chosen after a review of the group and online community literatures failed to reveal any study of individuals' pre-joining discussion volume expectations and preliminary runs indicated that model outcomes were not sensitive to variation in this parameter.

The final step in model calibration involved estimating the primary unobservable parameters by selecting values that provided the best fit between simulated samples of communities and the calibration sample. Participation cost (PC) and topic consistency cues (TCC) are central to the study's research questions, yet are not directly measurable with available data. To calibrate these parameters, we constructed measures of two observable community characteristics: member loss and discussion volume. Member loss was measured using overall proportional member loss. Daily membership loss was determined by comparing each day's membership list to the prior day's to determine how many individuals had left a community. Overall proportional member loss was then calculated by dividing the total number of members who left during the observation period by the total number of members present on the first day. Discussion volume was measured using average daily message volume, calculated by dividing the total number of community messages by the number of days in the observation period (130). These aggregate measures of emergent, but observable, community outcomes provided an empirical basis for characterizing communities that was less prone to noise due to seasonal effects or unobservable external factors.

We conducted a series of sessions, each simulating a sample of 96 communities, in order to calibrate participation cost and topic consistency cues. To best represent the initial state of the observed communities, all of which were well established, each simulation run consisted of two phases. In the first phase (initialization), 96 communities were created with initial sizes drawn from a distribution similar to that of the calibration sample (Log-Normal [μ = 4.18, σ = 1.56]). A 100-period initialization stage was run to represent each community's prior history. The length of this initialization period was selected based on pilot tests that indicated 100 was sufficient to move the simulated communities past the anomalous start-up stage, which was significantly more volatile than observed in established communities. The second phase (data collection) was 130 periods long, corresponding to the 130-day data collection period for the calibration sample. During the simulated data collection stage, membership lists and daily message totals were recorded. Measures of member loss and discussion volume were then constructed for each simulated community.

Between each calibration session, the unobservable parameters (PC and TCC) were varied to systematically explore the parameter space. When creating the samples of modeled communities, the parameters examined in the prior stages of calibration were drawn from the distributions described above (summarized in Appendix A). Participation cost was changed between sessions, varying between 0.1 and 1 in 0.05 increments. Similarly, values for topic consistency cues were varied between sessions from 0.01 to 0.1 in 0.01 increments. After each session, the mean and median of the member loss and discussion volume measures in the simulated sample of communities were compared with those observed in the calibration sample. Sample-level calibration was selected instead of community-level calibration to reduce the influence of idiosyncratic environmental factors that might affect the emergent characteristics of particular communities. We further refined the participation cost (PC) parameter by running additional calibration sessions in which PC was varied between 0.3 and 0.4 in 0.01 increments. Overall, the third and final stage of the calibration process resulted in identification of values for both PC (0.33) and TCC (0.02) that functioned as anchor points for subsequent validation and analysis.

Model Validation

During validation, the parameter distributions and values identified during calibration (Appendix A) are taken as given, the model is run, and the characteristics of the resulting communities are compared with empirical data in order to verify that the calibrated model is a reasonable representation of the phenomena of interest (Carley 1996). This analysis assesses how well, and under what conditions, the model represents the intended phenomena.

The model was validated using matched analysis (Law and Kelton 1991), a process involving simulation of specific cases and statistical comparison of simulation results and observed outcomes. A simulated community was created for each of the 96 communities in the validation sample. To create a matched pair of model and empirical results, the observable parameters—initial community size, attraction rate (A), and message contribution probability distribution (MCP)—were set based on data from a single community from the validation sample. The unobservable parameters—participation cost (PC) and topic consistency cues (TCC), initial expectations (F_0 and V_0), and interest distribution (I)—were set based on the results of model calibration (Appendix A). For each modeled community, PC and TCC were drawn once, and then 10 simulation runs were performed using these settings in order to have adequate statistical power (calculated to be

93.5%) (Van Voorhis and Morgan 2007). The proportional member loss and average daily message volumes were recorded. These measures of overall member loss and discussion volume were then averaged to create predicted values for each of the 96 communities.

Due to the non-normal distribution of community characteristics in the validation sample, the model results and empirical data were compared with the Wilcoxon signed rank test, a nonparametric paired sample test. Results indicated that emergent community characteristics predicted by the model were comparable to those seen in the empirical data. Member loss in the modeled and empirically observed communities were not significantly different ($p = 0.353$). Similarly, comparison of message volumes found no significant difference between the discussion volumes predicted by the calibrated model and those observed in the validation sample ($p = 0.224$).

While there were no significant differences between model outcomes and the empirical data, additional analysis of the validation results suggested that the model provides a more accurate representation for certain types of communities. Correlation analysis of member loss and discussion volume error,[10] indicates that the prediction error was lowest for communities with lower attraction rates, more active, concentrated participation (i.e., lower participation ratios and higher participation probabilities), and fewer initial members.[11] This may be a consequence of the approximations used to model member attraction and message contribution probabilities. Overall, validation indicates that the model provides reasonable predictions for communities in the range of sizes and discussion volumes commonly found in both online (Butler 2004) and traditional settings (McPherson 1983).

Analysis and Results

A virtual experiment was performed to analyze the model's predictions regarding the impact of alternative community platforms on community development. The experimental conditions were created by systematically varying parameters corresponding to platform attributes: participation cost (PC) and topic consistency cues (TCC). The experimental conditions included the values identified as most appropriate during calibration (PC = 0.33, TCC = 0.02) and incrementally higher and lower values to create 4 × 4 experimental structure (PC: 0.165, 0.33, 0.495, 0.66 and TCC: 0.005, 0.02, 0.035, 0.05). All other model parameters were set based on the values and distributions identified in calibration, an approach analogous to random assignment of subjects to different conditions.

For each condition, 500 communities were simulated for 100 initialization periods and 365 observation periods to ensure that sufficient statistical power was available to characterize the relationships between parameters and emergent outcomes. Model parameters and overall measures of size and resilience were recorded for each simulated community, which resulted in an analysis dataset of 8,000 simulated communities (N = 4 × 4 × 500). Community size was measured as the number of individuals present present after the final time period of each model run. Community resilience[12] was measured by calculating the mean fit expectation among members at the end of the final period in each simulation run.

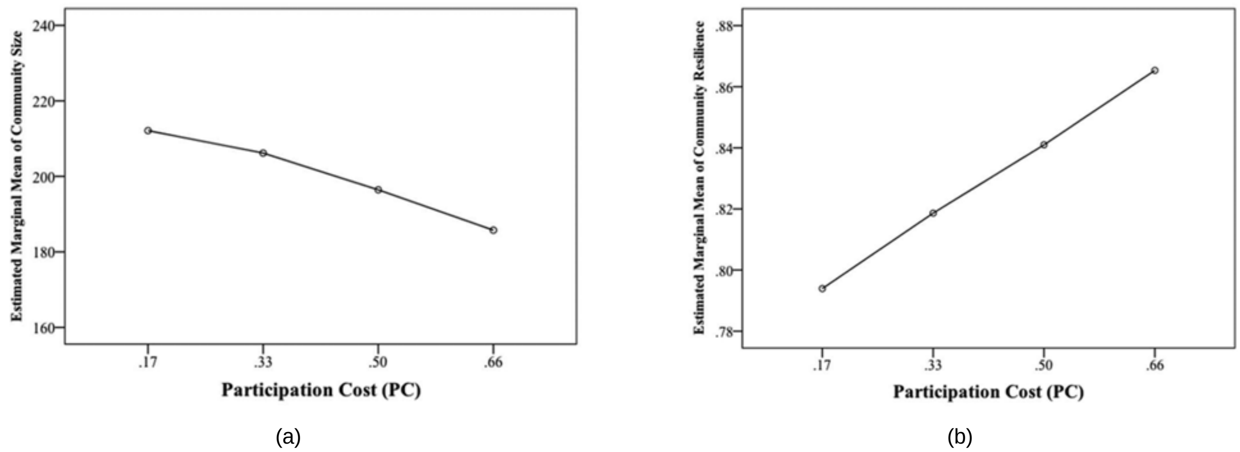
A between-subjects MANOVA was used to compare the effects of participation costs and topic consistency cues on emergent community characteristics. Community platform parameters (PC and TCC) were treated as fixed main effects, with community size and resilience as dependent variables. Other community attributes, including attraction rate, initial size, participation ratio, participation probability, and maximum interest range (Appendix A) were included as covariates. The corrected model was significant for both dependent variables (community size: adjusted $R^2 = 0.775$, $F(20,7979) = 1377.31$ and community resilience: adjusted $R^2 = 0.296$, $F(20,7979) = 168.82$).

[10]Absolute Error = Predicted Value – Observed Value.

[11]

Correlations (*p ≤ 0.5; **p < 0.01)	Member Loss Error	Discussion Volume Error
Initial Size	0.338**	0.290**
Attraction Rate	0.698**	0.646**
Participation Ratio	0.493**	0.415**
Participation Probability	-0.422**	-0.205*

[12]As an empirical validation of our measure of resilience, we tested whether it predicted member loss at future points in time and found that, consistent with our theorizing, mean fit was only moderately related to percentage loss of members. We used end-of-run measures and found a significant negative correlation (-0.156) between resilience/fit and percentage loss. An additional analysis of longitudinal data derived from the simulation found similar relationships for lagged models of resilience/fit_{t-1} and percentage member $loss_t$ (unstandardized Beta = -0.275, $p < 0.001$). In both datasets, we also examined the collective relationship between daily message volume, membership loss, and resilience/fit by testing a model in which resilience/fit is a mediator between communication activity and membership loss. The results support the argument that resilience/fit acts as a partial mediator between message activity and membership loss, further evidence that while it is clearly related to loss, resilience as operationalized here is not equivalent to it.



Note: Covariates appearing in the model are evaluated at the following values: attraction rate = 0.2958; participation ratio = 0.1606; participation probability = 0.0193; interest range = 0.4975; initial community size = 132.1750 (see Appendix A).

Figure 1. Results of the MANOVA Tests for Propositions 1 (Left) and 2 (Right)

Participation Costs

Both multivariate (Roy's largest root = 0.184; $p < 0.001$) and univariate ($F(2, 7979) = 19.06$; $p < 0.001$) tests of the impact of participation costs indicate a significant, negative relationship between participation costs and community size (Figure 1a). This implies that the OCASA model predicts that platforms designed to impose lower participation costs will lead to larger communities, because in those contexts individuals are willing to tolerate a higher signal to noise ratio (i.e., uninteresting messages/total messages). Higher message volumes associated with larger communities have a lower negative effect on size when participation costs are low because the amount of beneficial content necessary to maintain positive net benefit expectations is lower.[13] As a result, more members will be willing to remain involved in a community. Thus, the model predicts

> **Proposition 1**: Participation cost will be inversely associated with community size.

The MANOVA analysis also indicated a significant positive relationship between participation costs and mean fit expectations with a community ($F(2,7979) = 452.62$; $p \le 0.001$; see Figure 1b). This implies that platforms that impose higher participation costs will lead to communities that are more resilient. When participation costs are high, only those members whose interests align strongly with the community discussion will expect sufficient net benefits and, therefore, will remain. Higher expectations mean they are less affected by uninteresting messages; their expectations do not change as much when discussion content changes and they are, therefore, more willing to remain. Thus, the model predicts that a community with high participation costs will have a membership that, on average, is more resilient in the face of varying discussion topics.

Conversely, when participation costs are low, communities can attract members who are only marginally interested in a community's discussion topics and thus expect there to be a weak fit between their interests and the community discussion. However, because they have lower content fit expectations, their perceptions are more affected by the discussion activity, and as a result these marginal members will be more likely to depart if the current discussion happens to disproportionately emphasize topics that they find uninteresting. A community with lower participation costs, therefore, is expected to be less resilient because, on average, its members are more likely to leave in the face of even a small, temporary increase in uninteresting discussion activity. Thus the OCASA model predicts that

> **Proposition 2**: Participation costs will be positively associated with community resilience.

Topic Consistency Cues

Results of the virtual experiment indicate that topic consistency cues (TCC) are negatively related with community size ($F(2,7979) = 125.710$; $p < 0.001$; Figure 2). Community platforms with lower topic consistency cues reduce the impact a

[13] *Post hoc* regression analysis indicates that participation costs have a significant negative moderating effect on the relationship between discussion volume and member loss.

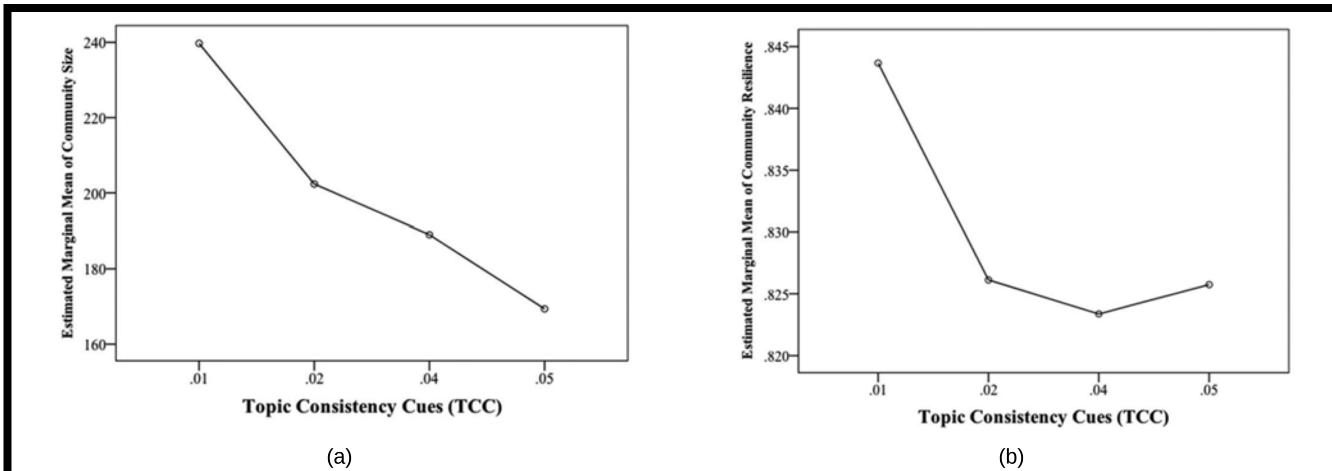


Figure 2. Results of the MANOVA Tests for Propositions 3 (Left) and 4 (Right)

message has on individual's expectations. In these communities, any particular message is interpreted as only a weak signal of what to expect in the future. While members will still adjust expectations based on what they observe lower TCC means that it takes more uninteresting messages to shift members from their initial, generally positive, expectations to such negative expectations that would cause them to leave. On average, when TCC is low more individuals remain involved, resulting in larger communities.

In contrast, community platforms that provide strong topic consistency cues increase the impact that messages have on individuals' expectations. Here, members are more likely to believe that current discussion topics will persist in the future. An observed message that is outside an individual's interest set, therefore, becomes a strong signal that future discussions will also be uninteresting. High levels of TCC lead individuals to assess expectations of future fit based largely on current discussion, resulting in content fit expectations that are less optimistic than their initial perceptions of a community. Low expectations of fit lead to lower expectations of net benefits, with more individuals choosing to leave, resulting in a smaller community. Thus, model predicts that

> **Proposition 3**: The strength of topic consistency cues will be inversely associated with community size.

The MANOVA analysis also indicates that community platforms with lower topic consistency cues are associated with more resilient communities ($F(2,7979) = 42.604$; $p \leq 0.001$). However, closer examination of the virtual experiment results (Figure 2) suggests a more complex relationship between topic consistency cues and community resilience. Low TCC results in communities with high levels of resilience by diminishing the impact of experienced discussion activity on individuals' expectations about the net benefit of future involvement. This allows optimistic initial expectations (Brinthaupt et al. 1991) to persist, even when the actual fit between discussion topics and individuals' interests is low. Thus, communities with low TCC will tend, on average, to have members with higher content fit expectations, and hence be more resilient.

On the other extreme, when strong topic consistency cues lead members to believe each message is highly indicative of future discussion topics, their expectations change more abruptly. High TCC magnifies the impact of experienced discussion activity on individuals' expectations about future content fit. The tight coupling of future fit expectations with current discussion activities has two complementary effects: it increases the likelihood that marginally interested members will filter out of the community, and it reinforces the expectations of highly interested members. When individuals' assessment of the net benefit of continued involvement, and hence their likelihood of attrition, is highly influenced by the idiosyncrasies of particular batches of discussion, they are more likely to conclude they should leave a community after encountering a single batch of disproportionately uninteresting messages. This tendency results in attrition of those who are not interested in a majority of topics that have been discussed within a community. At the same time, those that remain are most likely to be interested in the current topics, and hence will be likely to contribute messages on those topics in the future, increasing the degree to which current activity is actually predictive of future fit. Together these effects result in individuals who are in communities with high TCC having, on average, higher content fit expectations—that is, being more resilient.

However, the results of the virtual experiment (Figure 2) suggest that platforms with moderate levels of topic consistency cues are expected to lead to lower levels of community resilience. In these communities, the influence of changing message topics on members' expectation of net benefits is strong enough to lower their initial optimistic expectations. This increases the number of members who have low expectations for their future involvement and thereby reduces community resilience. At the same time, in moderate TCC communities these effects are not strong enough for the filtering and reinforcement effects of TCC to drive up the average expectations. As a result, moderate TCC communities are expected to have lower community resilience. Taken together, the model predicts that

> **Proposition 4**: The relationship between topic consistency cues and community resilience is curvilinear. Community platforms with very low and very high topic consistency cues are associated with greater community resilience; platforms that signal moderate topic consistency cues are associated with lower community resilience.

Discussion

The proposed ASA theory of online communities provides a basis for understanding the emergence of community size and resilience from the interplay of members' participation decisions and community technology characteristics. By extending ASA theory, this study theorizes how technology-related aspects of community platforms—participation cost and topic consistency cues—affect community-level outcomes. In what follows, we discuss a synthesis of the model's findings and the implications arising from them.

Taken together, our results have important implications for understanding changes in community technologies. Community leaders who face community development challenges tend to pursue the implementation of "new" or "missing" features in hopes of improving the value and experience they provide for individual participant (Sharma et al. 2011). While our goal was not to provide a detailed examination of such features, analysis of the OCASA model suggests that when altering a community platform it is also critical to consider how changes might affect emergent community characteristics such as size and resilience. Proposition 1 suggests that careful choices among features that raise or lower participation costs can be a powerful strategy for affecting the emergent size of an online discussion community. Characterizing common community platform features in terms of their likely impact on participation costs (Table 4) can, therefore, assist leaders in managing, and anticipating, the impact of their choices on a community.

When community administrators want to alter community size, technologies that affect participation costs are likely to provide a powerful leverage point. In general, larger communities are able to generate more benefits for members (e.g., more resources from which to draw) and to attract more advertisers (e.g., generate larger revenue possibilities). But some administrators, such as those working with health and emotional support groups, may want to maintain smaller communities to support to the goals of the membership. The results of the model analysis suggest that changing community platforms in ways that raise (or lower) the participation costs incurred by members is likely to have a significant impact—whether it is intended or not.

However, Proposition 2 suggests that reducing participation costs may have more nuanced effects beyond simply increasing community size. Community administrators who implement new technologies in order to lower participation costs must consider that this may also decrease community resilience. By reducing the degree of fit members require to maintain continued participation, technologies that lower participation costs primarily change the assessments of members whose interests only marginally match a community's discussion activity. With lower participation costs, individuals with a weak alignment between community discussions and their interests will be more likely to remain in the community, increasing community size but decreasing overall community resilience. Conversely, technologies that impose higher participation costs cause members to depart when their interests do not align well with community discussions, resulting in a more resilient membership that is less affected by messages that they perceive as "noise."

Although communities create benefit through member-generated discussion, it is the expectation of future benefits that retains current members and attracts new members (Hagel and Armstrong 1997). The focus of potentially beneficial conversations (Arguello et al. 2006) emerges from the aggregated interests and contribution tendencies of the members, features of the community that are both largely invisible and inherently dynamic (Baym 2000). Individuals thus face the challenge of developing expectations about the future benefits of involvement with, at best, incomplete knowledge of the current and future interests of the community. Topic consistency cues may be influenced by technological features (Table 5) that help members decide whether current discussion activity is a good indicator of the interests of current and future community members, and hence whether it should be treated as a strong signal of future discussions and the associated benefits.

Table 4. Technology Features that Impact Participation Costs	
Decrease	*Search*: Enables users to find messages of interest based on specific keywords or terms provided, thereby reducing time spent reading irrelevant messages.
	Subscriptions: Automatically notifies users when messages that match their interests are posted, which reduces time finding interesting content and reading uninteresting content.
	Threaded forum: Messages are organized by discussion topic, allowing users to skip uninteresting topics.
	Message Display Options: Sorting messages by topic lets users identify discussions likely to be of interest. Indicating the number of replies helps users identify active discussions. Sorting messages by recency lets users skip over older messages.
	Thread Preview on Mouse Over: Provides a glimpse of message detail that reduces the number of clicks (and time) necessary to decide if a thread is worth reading.
	Unread Message Tracking: Identifies messages a user has not yet seen, reducing the time spent unintentionally re-reading messages.
	Relevant Thread Identification: When users type to create a new thread, software brings up existing threads that may be a match, increasing the grouping of message in threads.
	CAPTCHA: Requires a human to enter a visually scrambled code to restrict access for "bots," reducing message volume and uninteresting messages generated by spam programs.
Increase	*User Registration for Access*: Creates barriers of entry to accessing messages, thereby increasing costs.
	Advanced search: Requires additional parameters to be input prior executing a search that can increase the time to perform simple searches.
	Digest messages: Automatically sends users all new messages for a time period; however, messages are not organized by interest or topic, which can result in a large number of uninteresting messages a user must process to find messages of interest.
	Limited searches (time): Limits how frequently searches can be performed (e.g., 1 every 10 seconds) makes searching for interesting messages more time consuming.
	Limited searches (site): Llimits results to particular part of the community can make it harder to find content across the community (requiring multiple searches)

Table 5. Technology Features that Impact Topic Consistency Cues	
Decrease	*"Flat" forum*: A new message is added to end of a discussion, with no relation to any prior message.
Increase	*Threaded forum*: Messages are organized by topic, providing members cues to the more persistent topics of interest to a community.
	"Sticky" Messages: Messages that are permanently and prominently placed in an area where users can easily see sends signals as to what the community finds interesting.
	Similar Thread Displayed: Indicates the presence other related threads.
	Thread View Count: Shows number of times a message has been viewed, and thus is a proxy of what the community members find interesting.
	Thread Reply Count: Shows the number of times members have replied to a message, which is an indicator of topics that are of interest to community members.
	Post Rating: Members can provide ratings of messages (e.g., thumbs-up/thumbs-down, stars)
	Hot Topic Icon: Indicates messages that are currently receiving an extraordinary amount of attention, signaling this is what community members find interesting.
	Member Profile: Member information attached to messages. This allows readers to assess the prominence of the contributor in the community discussion, and how likely they (and topics they are interested in) are to reoccur.
	Subforums: An organizational method that groups messages based on topics. The size and activity level signals the presence of community members who find these topics interesting.

Individuals are necessarily optimistic about the benefits of involvement when they join an online community. While the degree may vary, all new members believe there will be some level of fit between their interests and the community's discussion activity. But these initial expectations change as they encounter the real discussion. Community platforms that provide more information about the composition and interests of members increase topic consistency cues (Table 5), magnify the impact of current discussions on individuals' expectations for future benefits, and affect community size (Proposition 3) and resilience (Proposition 4). These results suggest that technologies that help members feel confident about what topics are likely to be discussed in the future can have unexpected and potentially undesirable consequences for the community as a whole by reducing community size (P3) and community resilience (P4) in some situations.

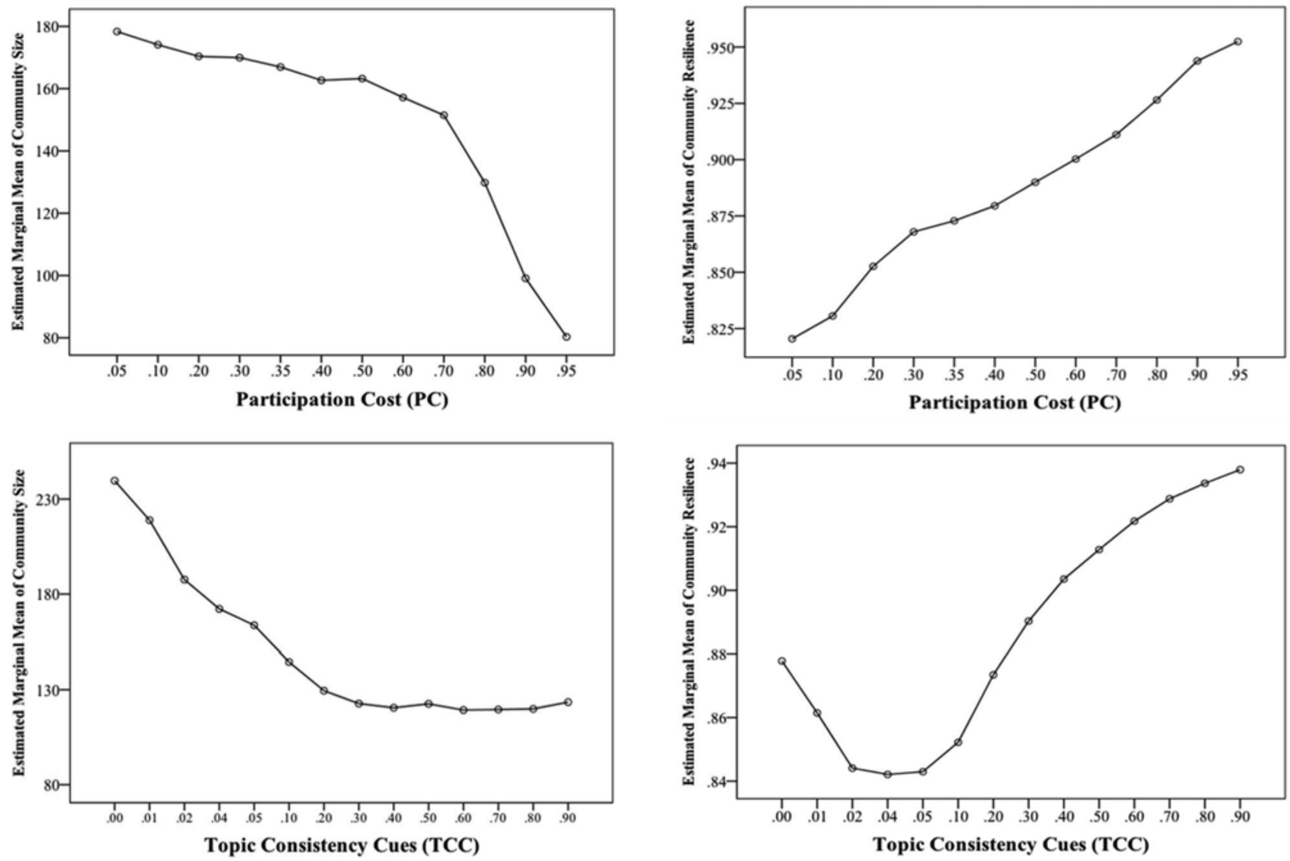
It is worthwhile to consider how our model would perform if communities used technologies that produced significantly higher, or lower, PC and TCC values. Recall that the calibrated values for PC and TCC were extracted from a population of communities based on a single underlying technology, which provided an empirically grounded anchor point for the analysis presented above. To better understand the behavior of the model, we also conducted analyses that considered the effect of PC and TCC on community size and resilience across the full range of possible parameter values. In this supplemental analysis, PC ranged from 0.05 to 0.95 (compared with 0.165 to 0.66 around a calibrated value of 0.33) and TCC ranged from 0.001 to 0.9 (compared with a range of 0.005 to 0.05 around a calibrated value of 0.02). The results of the extended analyses (Figure 3) are consistent with those of the more conservative analyses presented earlier (Figures 1 and 2). Even when simulation values for TCC extend beyond our calibration range, trends in community size and resilience are consistent with the propositions we derived above.

Notably, the curvilinear relationship between TCC and community resilience is much more pronounced than in the original. While the relationship between participation cost (PC) and community size is generally negative, it takes a significantly greater downward slope for values of PC $\geq$ 0.7. Similarly, the relationship between TCC and community size remains negative far beyond the calibrated range, but its effect seems to level off for values of TCC greater than 0.3. Although the exact impact of the specific technology features identified in Table 4 and Table 5 on PC and TCC cannot be known in advance, this extended analysis suggests the model can be expected to have reasonable and testable predictions for a wide range of possible platform arrangements. With the caveat that the model is likely to be less precise for parameter values that are further from the empirically calibrated values (as found in the validation with the observable parameters), these results suggest that online communities may be subject to different community-level dynamics depending on the nature of the underlying community platforms.

Our analysis suggests several ways in which developers might inadvertently undermine their own efforts to grow and strengthen an online discussion community. First, the results suggest that many "obvious improvements" may have unanticipated consequences. From a systems design perspective, reducing participation costs and providing additional information about community activity sound like good things to do. Increasing ease of use, improving usability, and lowering participation costs are all generally positive design goals. However, the OCASA model indicates that in the context of a dynamic community with heterogeneous individuals, even these straightforward "improvements" can have unexpected, even undesirable, consequences for the community as a whole, possibly producing a less resilient member base (P2, P4) and reducing community size (P3).

A more subtle implication of these results is that typical interface and system design processes may need to be altered when developing online community platforms. Traditional user-oriented design practices use surveys, interviews, and focus groups to gather information about users' responses to potential design changes. Individuals with strong expectations of future benefit from the community are more likely to contribute to the health and success of the community by participating in these information-gathering activities. However, our results suggest that technology changes that appeal to these individuals are not likely to either significantly impact their own participation in the community or produce the community-level outcomes that their perceptions would lead them to expect. When individuals' responses to technology features depend on their degree of expected fit, making technology design decisions based on feedback from highly involved members in likely to have unanticipated—and potentially undesirable—implications for longer-term success. That is, they may actually reduce community resilience (P2, P4) or reduce community size (P3) through their best intentions, improving the experience for individual users by lowering participation costs and/or altering topic consistency cues.

Additional *post hoc* analysis of the model results suggests that participation costs also interact with the breadth of interests in a community's target membership, which affects the nature and difficulty of the design problem associated with creating a sustainable community. When community technologies have high participation costs, there is little room for design experimentation because members leave if they see even a



Note: Covariates appearing in the model are evaluated at the following values: attraction rate = 0.2974; participation ratio = 0.1601; participation probability = 0.0192; interest range = 0.4998; initial community size = 132.8572 (see Appendix A).

Figure 3. Results for Community Size and Resilience from Extended PC and TCC Range

few messages that do not match their interests; under these conditions, establishing a sizable community can be quite difficult. If participation costs are high, community sustainability can be increased by narrowing the topical interest focus of the community and targeting individuals for whom that entire focused topic set is of interest. In this scenario, high participation costs serve as a disincentive to the marginally interested member, which keeps content relevant and consistent with the interests of the primary targeted participants. Of course, this approach is dependent on the existence of a sufficient number of individuals with that particular interest. If the topic is narrowed to the point where there is insufficient interest among potential participants, a community will not be sustainable. In sum, when participation costs are high, community sustainability is highly dependent on the ability to identify a topical focus that is of interest to enough people that there is sufficient discussion activity, but not narrow enough that participants are likely to be interested in the entire topic set.

As participation costs drop, the design problem facing community developers changes, with the greatest reduction in community size occurring when moving from high to medium cost conditions. Reducing participation costs can alter the behavior of individuals who are tangentially interested and have uncertain expectations. This, in turn, can significantly change the content and volume of discussion activity and thus the nature of a community. This shift is most clearly seen in communities with low participation costs and a broad topic set (lower left corner of Figure 4): they attract a large number of members, but their membership has low resilience. When the population of potential participants is large, low resilience is less of a problem because "replacement" members are often available. However, when the target population is small, as is the case with many organizationally affiliated communities, the combination of large size and low resilience is unlikely to be sustainable. Under these conditions, developers seeking to create sustainable communities should focus less on reducing participation costs and more on identifying a topical focus that

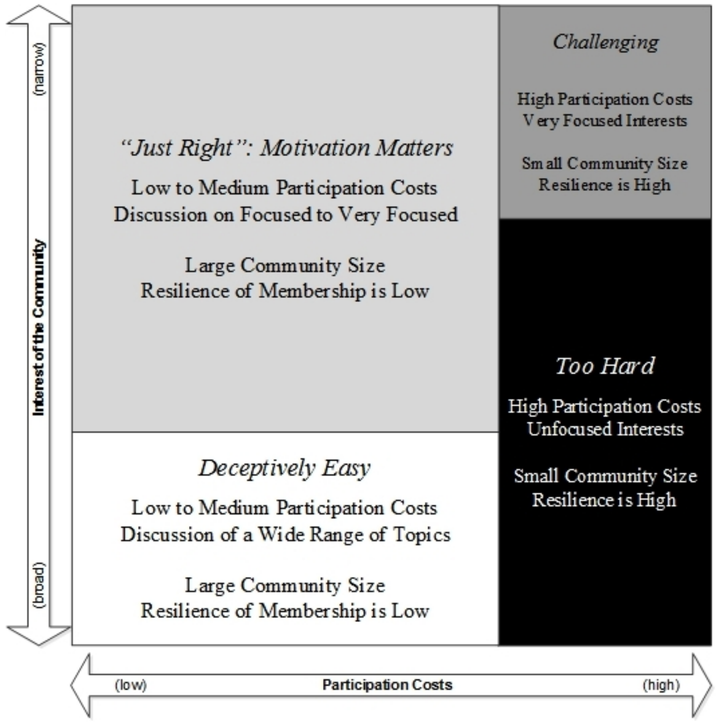


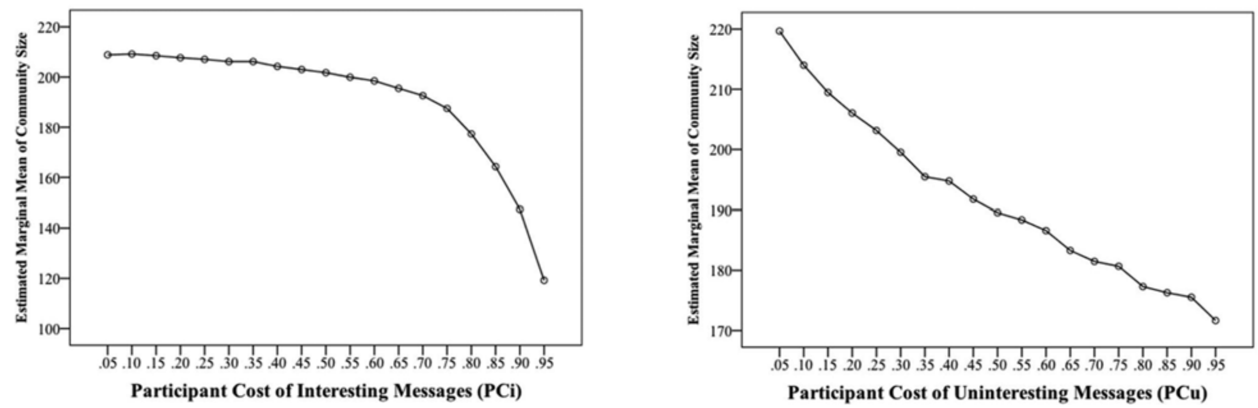
Figure 4. Participation Costs and Interest Focus and Community Sustainability

is a strong fit with a significant proportion of the target population. Taken together these *post hoc* results suggest that developers should employ a Goldilocks principle[14] of community definition, selecting topical boundaries that are broad enough to attract sufficient members, but narrow enough to allow for good fit between individuals' interests and the aggregate activities of the community.

Another way that the OCASA model can be used is as the basis for variant models that examine the potential impacts of alternative community platform designs. For example, some community platforms extend basic messaging capabilities by reducing the costs that members incur when dealing with messages that do not interest them. Whether by clustering related messages or through more sophisticated prioritization and collaborative filtering capabilities, such community features impose differential participation costs depending on whether a member finds (or is likely to find) a message interesting. To examine the implications of this alternative platform design, the foundational OCASA model was modified to distinguish between participation costs imposed by messages that an individual finds interesting (PC_i) and those the individual finds uninteresting (PC_u). Analysis over the full range of possible PC_i and PC_u values suggests that the implications for community size and resilience differ for these two aspects of participation cost (Figures 5 and 6).

The MANOVA results for the analysis of this model indicate that both PC_i and PC_u are significantly related to community size and resilience. Examination of these relationships suggests a more nuanced community-level strategy for managing participation costs. When participation costs associated with interesting messages are high relative to the benefit that members receive (i.e., $PC_i > 0.7$), lowering these costs will result in a community that is larger and less resilient, as shown in Figures 5(a) and 6(a), consistent with findings from the foundational model captured in Propositions 1 and 2. However, if PC_i is lower ($PC_i < 0.7$), then the community level impact of reducing it further is much less. In contrast, lowering the costs associated with uninteresting messages (PC_u) produces a linear increase in community size—Figure 5(b)—and a negative impact on community resilience that significantly increases for values of $PC_u < 0.3$, as shown in Figure 6(b).

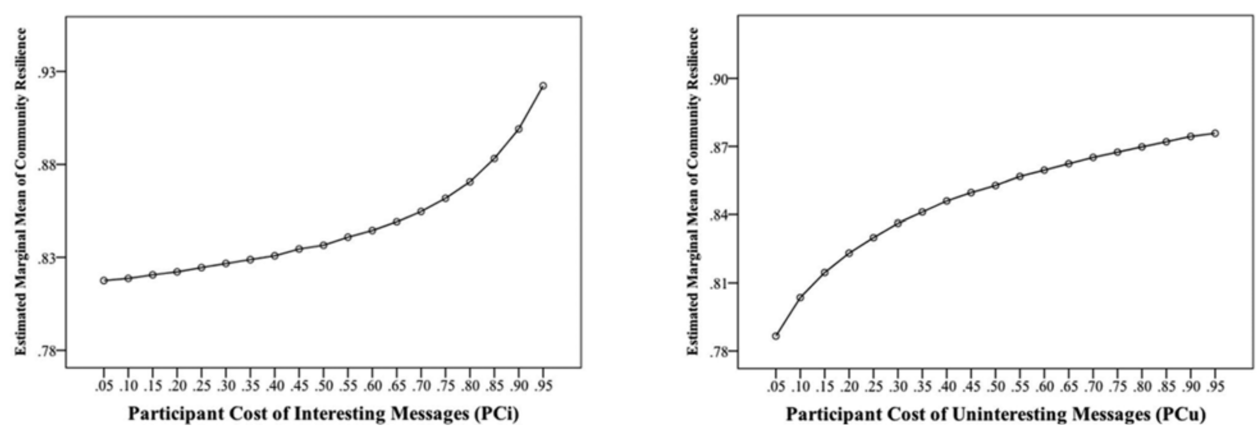
[14]The Goldilocks principle states that conditions must fall within certain margins rather than reaching extremes. Closely related to the theory of equilibrium, it is an ideology and not a logical principle. The Goldilocks principle has been applied across many disciplines, including biology, astrobiology, healthcare, and economics.



(a) (b)

Note: 5(a) at mean levels of PC_u, 5(b) at mean levels of PC_i. Covariates appearing in the model are evaluated at the following values: attraction rate = 0.2941; participation ratio = 0.1599; participation probability = 0.0193; interest range = 0.5000; initial community size = 134.7485 (see Appendix A).

Figure 5. Implications of Differential Participation Costs (PC_i and PC_u) for Community Size



(a) (b)

Note: Covariates appearing in the model are evaluated at the following values: attraction rate = 0.2958; participation ratio = 0.1606; participation probability = 0.0193; interest range = 0.4975; initial community size = 132.1750 (see Appendix A).

Figure 6. Implications of Differential Participation Costs (PC_i and PC_u) for Community Resilience

Together these analyses suggest that communities seeking to balance community size and resilience would be well served by platforms that impose high participation costs for uninteresting messages. By reducing PC_u from high values, community size could still be increased with little negative impact on resilience, but reductions from lower PC_u values result in significantly lower community resilience without providing a proportionately greater increase in community size. Further, the model suggests that once the costs associated with interesting messages (PC_i) are below a relatively high threshold (0.7) there is little consequence of further reduction, at least with respect to community size and resilience. Community designers and administrators would, therefore, benefit most from designing platforms that impose lower participation costs for uninteresting messages, and largely ignore the costs associated with messages individuals find interesting. While this is consistent with a casual reading of practitioner recommendations and conversations with community managers, future work could empirically examine the predictions of this, and other, variant OCASA models.

Table 6. Participation Costs and Topic Consistency Cues and Social Media Technologies

Media	Participation Cost Influence	Media	Topic Consistence Cue Influence
	News Feeds – Provides push notifications and summaries of activities by those in a members network, making it easier for to find current content that might be of interest		*Timeline* – Organizes all profile content of a user, making it more easy for users to understand the current interests of others
	Ticker – Shows all friend and page updates in one place as soon as they are posted, making them more discoverable, allowing users to easily see content of interest to them		*Likes* – Shows the number of times a message/post/page has been clicked as being liked by users, and thus is a proxy of what the community finds interesting
	"Tweet" – A 140 character message, which reduces time required to create and post		*Trending Topic* – An algorithm identifies topics that are immediately popular, helping users discover the hottest emerging topics of discussion on the site
	Following – The ability to have tweets from a particular member pushed to users, making it easier to keep up with members one finds to be of interest		
	Hashtag (#) – Convention used to identify topic of message, which facilitates searching for tweets on the topic		*Hashtag (#)* – Convention used to group posts together by topic, providing users cues to topic of the message
	Updates – Digest of changes or comments relevant to users' networks, making it easier for them to find content relevant to them		*Social Bookmarking* – Facilitates users in collecting and categorizing content, making it easier to identify what content is about
	Circles – Allows member to identify a sub-set of others to communicate, making it more efficient to communicate with those they find interesting		*Social Filtering* – Content posted by members is voted "up" or "down" by the community, signifying content that is of interest to the community
	RSS: Automatically pushed to users content they are interested based on their subscription, which reduces time finding interesting content and processing uninteresting content		*Social Curation* – Tools facilitate the collection of media around specific interest, making it easier to identify what content is about

Implications for Social Media

Although the OCASA model is rooted in prior studies of online discussion communities, our findings also have implications for theorizing and designing other social media. They may seem to be radical departures from prior systems, but at their conceptual core social media are simply a continued evolution of group-based computer-mediated communication platforms. Social media provide similar core capabilities as their technological predecessors, facilitating communication among individuals who have shared interests. Similar socio-psychological processes may, therefore, be expected to drive participation (i.e., evaluations of net benefits based on attraction, selection, and attrition). Indeed, participation costs and content consistency cues apply equally well in characterizing many of the popular features associated with leading social media technologies (Table 6). To the extent that these features positively or negatively influence PC or TCC, their impact on community size and resilience would be reflected in our model.

The OCASA theory and the simulation results presented here may also shed new light on some recent social media phenomena. For example, Myspace was for some time the dominant leader in social networking, finding success by offering users a relatively low-cost platform for participating in many socially oriented discussion communities. This strategy supported rapid growth, but may have resulted in a less resilient membership that was not willing to remain when the activity of the community shifted to different topics. In contrast, its successor, Facebook, has repeatedly demonstrated that, with careful management, it is possible to make technological changes that increase participation costs (e.g., news feed, timeline), even to the dismay of highly involved participants. Of course, network effects also have played a role in Facebook's success, but often lost in the discussion is the reality that social media network effects are instantiated through technology use; if nobody actually communicates, network effects cannot come to play. The OCASA model suggests that higher participation costs can serve to strengthen a community. When coupled with a willingness to ignore feedback from those individuals who already have a strong positive expectation of future benefits, this might in part explain why Facebook has been able to develop a more resilient membership. Understanding how community platforms mediate the relationship between potential benefits, realized benefits, and perceptions of expected benefits remains an important focus for IS research, and one that the work presented here may begin to illuminate.

Limitations and Directions for Future Research

As with all theory development studies, this simulation should be seen in terms of a balance between purpose, the relationship of the model characteristics to that purpose, and the use of empirical data relative to that purpose (Burton and Obel 1995). Within this scope, we note several limitations of our approach.

First, in following guidelines of simulation that recommend focusing on a few key constructs, the characteristics we included served their purpose of strategically extending ASA theory into the context of online communities and exploring the interplay of individual expectations and community characteristics. Here, we explicitly chose not to theorize specific technological antecedents of PC and TCC; a range of technological features could impact PC and TCC (as suggested in Tables 4 and 5), but for our purposes there was little benefit to validating the extent of their impacts on PC and/or TCC. While extending the model to include specific technology features might increase its "realism," it goes beyond our purpose of developing a general model that could apply to a broad range of specific technologies. However, future researchers could build on our findings by empirically documenting how specific technology features affect PC and TCC and creating extensions of the OCASA model based on those results.

Second, our assumption was that a particular cost–benefit logic underlies individuals' participation decisions, but this transactional approach is only one of several possible ways of conceptualizing individual choice (Ridings and Gefen 2004). Specifically, because we operationalize PC on a per-message basis that is consistent for all participants in a community, the model may not accurately characterize the dynamics in communities that are subject to systematic individual variation in participation costs and benefits. Obtaining benefits in proportion to the number of messages read seems appropriate in topic-based communities, such as technical support groups, but it may be less accurate in a bond-based community (Ren et al. 2012). Although participants in bond-based communities still obtain benefits and incur costs, additional factors may affect participation decisions (e.g., message quality, community responsiveness, and personal visibility). Members may also act altruistically, for the good of the community (Bateman, Gray, and Butler 2011; Wasko and Faraj 2005) in ways that are not captured in a cost–benefit model, but we do not consider these more complex individual-community relationship motivations in our model. The extent to which such other factors drive member expectations and participation decisions in ways that are inconsistent with cost–benefit logic is a limit on the generalizability of the simulation model presented here. Future research that develops, tests, and analyzes alternative operationalizations of the OCASA theory would strengthen its value as a foundational theory for describing the role of community platform technologies in community growth and sustainability.

Finally, the data used to calibrate and validate the simulation model came from push-based communities. However, we are not theorizing the push nature of the community platform, but rather, theorizing communities in terms of fit assessments, participation costs, and topic consistency cues. While our analyses are therefore relevant to discussion-based communities in general, future work could extend the model to theorize other differences in form (e.g., push versus pull, synchronous versus asynchronous) and social structure (e.g., social networks) that may impact communication. Individuals might also experience different PC and TCC because of their own skills, experiences, and cognitive capabilities. While we did not model such individual level variation, future models might productively expand on our work by elaborating the operationalization of the proposed OCASA theory in these ways and collecting community-level data from a broader array of community platforms in order to calibrate and test more elaborate simulation models.

Conclusion

Online communities are different from traditional information systems; they are emergent sociotechnical systems whose design and operation requires an understanding of multilevel social processes. One prominent characterization of the IS design–outcome relationship distinguishes between three types of IS: functional, enterprise, and network (McAfee 2006). The impact of functional IS lies in increasing individual productivity (Robey and Boudreau 1999), while enterprise IS positively impact outcomes through improved business processes, and network IS create value by supporting unstructured peer-to-peer interactions. As a kind of networked IS, online discussion communities are inherently evolutionary, as technology choices and user behaviors interact to shape their emergent character, resources, and capabilities (Thompson 2005). Methodologies for the design of functional and enterprise systems, such as user needs analysis and business process analyses (Sabherwal and Robey 1993), are, therefore, ill-suited to the design and development of online communities (Zammuto et al. 2007).

Because online communities are fundamentally emergent systems (e.g., Fulk 1993; Walther 1996), researchers need theories to help understand the dynamic interplay between individuals and community technologies. A useful online community theory should articulate the value of key community characteristics, outline a way of describing different technology choices, and explain how technology characteristics interact with the other aspects of a community and give rise to particular outcomes (Venable 2006). Prior research has focused primarily on understanding how community characteristics lead individuals to choose to participate (or not), but has only minimally addressed the dynamic aspects of online community development. Although the immediate effects of new technologies on mechanistic efficiency may be most visible, it is incorrect to assume that these first-order effects are necessarily the most important consequences (Sproull and Kiesler 1991). Ultimately, how community members behave and how communities function are the result of the complex interaction of technology choices, others' behaviors, and emergent community characteristics.

Our integrative theoretical approach extends the literature that seeks to understand the dynamic nature of online communities. For example, the resource-based model of online communities (Butler 2001) frames community development and sustainability as a cyclical process linking resources and benefits. While this approach explains *why* online communities may survive, it is less useful as a theory to those interested in *how* community platforms affect a community's ability to survive (Venable 2006). By theorizing the underlying constructs and causal mechanisms that can be impacted by technology choices, and combining empirical data with computational modeling, this study provides a foundation for future research that tests and extends OCASA to build more robust, integrative models of online community dynamics and development.

Acknowledgments

This research was partially funding by the National Science Foundation, grants CSS9711548 and OCI-0951630, the McIntire School of Commerce's Montague Endowment for Faculty Research, and the Deloitte & Touche Faculty Research Fund.

We would also like to thank the senior editor, Suprateek Sarker, the associate editor, and the entire team of reviewers, including the consulting reviewer from the Operations Management discipline, for their contributions in guiding our paper to acceptance.

References

Andrews, D. C. 2002. "Audience-Specific Online Community Design," *Communications of the ACM* (45:4), pp. 64-68.

Arguello, J., Butler, B., Joyce, E., Kraut, R., Ling, K. S., and Wang, X. 2006. "Talk to Me: Foundations for Successful Individual-Group Interactions in Online Communities," in *CHI 2006: Proceedings of the ACM Conference on Human Factors in Computing Systems*, New York: ACM Press, pp. 959-968.

Attaway, S. 2012. *MATLAB: A Practical Introduction to Programming and Problem Solving* (2nd ed.), Waltham, MA: Elsevier Inc.

Bagozzi, R. P., and Dholakia, U. M. 2006. "Antecedents and Purchase Consequences of Customer Participation in Small Group Brand Communities," *International Journal of Research in Marketing* (23:1), pp. 45-61.

Baldwin, C., and von Hippel, E. 2011. "Modeling a Paradigm Shift: From Producer Innovation to User and Open Collaborative Innovation," *Organization Science* (22:6), pp. 1399-1417.

Bapna, R., Goes, P., and Gupta, A. 2003. "Replicating Online Yankee Auctions to Analyze Auctioneers' and Bidders' Strategies," *Information Systems Research* (14:3), pp. 244-268.

Bateman, P. J., Gray, P. H., and Butler, B. S. 2010. "Should I Stay or Should I Go: The Role of Referrals on Online Community Member Turnover and Tenure," in *Proceedings of the 31st International Conference of Information Systems*, Saint Louis, Missour, December 12-15.

Bateman, P. J., Gray, P. H., and Butler, B. S. 2011. "The Impact of Community Commitment on Participation in Online Communities," *Information Systems Research* (22:4), pp. 841-854.

Bateman, P., Pike, J. C., and Butler, B. S. 2011. "To Disclose or Not: Publicness in Social Networking Sites," *Information, Technology & People* (24:1), pp. 78-100.

Baym, N. K. 2000. *Tune in, Log On: Soaps, Fandom, and Online Community*, Thousand Oaks, CA: Sage Publications.

Beenen, G., Ling, K., Wang, X., Chang, K., Frankowski, D., Resnick, P., and Kraut, R. 2004. "Using Social Psychology to Motivate Contributions to Online Communities," in *Proceedings of the 2004 Conference on Computer Supported Cooperative Work*, New York: ACM PRess, pp. 212-221.

Benbasat, I., and Zmud, R. W. 2003. "The Identity Crisis Within the IS Discipline: Defining and Communicating the Discipline's Core Properties," *MIS Quarterly* (27:2), pp. 183-194.

Blanchard, A. L., and Markus, M. L. 2004. "The Experienced 'Sense' of a Virtual Community: Characteristics and Processes," *The Data Base for Advanced Information Systems* (35:1), pp. 64-79.

Boyd, D., and Ellison, N. 2007. "Social Network Sites: Definition, History, and Scholarship," *Journal of Computed-Mediated Communication* (13:1), pp. 210-230.

Bratley, P., Fox, B., and Schrge, L. 1987. *A Guide to Simulation* (2nd ed.), New York: Springer-Verlag.

Bretz Jr., R. D., Ash, R. A., and Dreher, G. F. 1989. "Do the People Make the Place? An Examination of the Attraction–Selection–Attrition Hypothesis," *Personnel Psychology* (42:3), pp. 561-581.

Brinthaupt, T. M., Moreland, R. L., and Levine, J. M. 1991. "Sources of Optimism among Prospective Group Members," *Personality and Social Psychology Bulletin* (17:1), pp. 36-43.

Burke, M., Marlow, C., and Lento, T. 2009. "Feed Me: Motivating Newcomer Contribution in Social Network Sites," in *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, New York: ACM Press, pp. 945-954.

Burton, R. M., and Obel, B. 1995. "The Validity of Computational Models in Organizational Science: From Model Realism to Purpose of the Model," *Computational and Mathematical Organization Theory* (1:1), pp. 57-71.

Butler, B. S. 2001. "Membership Size, Communication Activity, and Sustainability: The Internal Dynamics of Networked Social Structures," *Information Systems Research* (12:4), pp. 346-362.

Butler, B. S. 2004. "When Is a Group Not a Group: An Empirical Examination of Metaphors for Online Social Structure," in *Proceedings of the Academy of Management Annual Meeting*, New Orleans, LA.

Butler, B. S., Sproull, L., Kiesler, S., and Kraut, R. 2007. "Community Effort in Online Groups: Who Does the Work and Why?," in *Leadership at a Distance,* S. Weisband and L. Atwater (eds.), Mahwah, NJ: Lawrence Erlbaum Associates.

Cable, D. M., and Judge, T. A. 1996. "Person-Organization Fit, Job Choice Decisions, and Organizational Entry," *Organizational Behavior and Human Decision Processes* (67:3), pp. 294-311.

Carley, K. M. 1991. "A Theory of Group Stability," *American Sociological Review* (56:3), pp. 331-354.

Carley, K. M. 1996. "Validating Computational Models," unpublished paper, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA.

Carley, K. M. 2001. "Computational Approaches to Sociological Theorizing," in *Handbook of Sociological Theory,* J. Turner (ed.), New York: Kluwer Academic/Plenum, pp. 69-84.

Carroll, G., and Harrison, J. R. 1998. "Organizational Demography and Culture: Insights from a Formal Model and Simulation," *Administrative Science Quarterly* (43), pp. 637-667.

Chen, A. N. K., and Edgington, T. M. 2005. "Assessing Value in Organizational Knowledge Creation: Considerations for Knowledge Workers," *MIS Quarterly* (29:2), pp. 279-309.

Choi, B. R., Alexander, K., Kraut, R. E., and Levine, J. M. 2010. "Socialization Tactics in Wikipedia and Their Effects," in *Proceedings of the 2010 ACM Conference on Computer-Supported Cooperative Work,* New York: ACM Press, pp. 107-116.

Choua, S. -W., Min, H. -T., Chang, Y. -C., and Lin, C. -T. 2010. "Understanding Continuance Intention of Knowledge Creation Using Extended Expectation-Confirmation Theory: An Empirical Study of Taiwan and China Online Communities," *Behaviour & Information Technology* (29:6), pp. 557-570.

Cohen, M. D., March, J., and Olsen, J. P. 1972. "A Garbage Can Model of Organizational Choice," *Administrative Science Quarterly* (17), pp. 1-25.

Constant, D., Sproull, L., and Kiesler, S. 1996. "The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice," *Organization Science* (7), pp. 119-135.

Cummings, J. N., Butler, B., and Kraut, R. 2002. "The Quality of Online Social Relationships," *Communications of the ACM* (45:7), pp. 103-108

Davis, J. P., Eisenhardt, K. M., and Bingham, C. B. 2007. "Developing Theory through Simulation Methods," *Academy of Management Review* (32:2), pp. 480-499.

Dennis, A. R., Wixom, B. H., and Vandenberg, R. J. 2001. "Understanding Fit and Appropriation Effects in Group Support Systems Via Meta-Analysis," *MIS Quarterly* (25:2), pp. 167-193.

Donath, J., Karahalios, K., and Viégas, F. 1999. "Visualizing Conversation," *Journal of Computer-Mediated-Communication* (4:4).

Ducheneaut, N. 2005. "Socialization in an Open Source Software Community: A Socio-Technical Analysis," *Computer Supported Cooperative Work* (14:4), pp. 323-368.

Faraj, S., Jarvenpaa, S. L., and Majchrzak, A. 2011. "Knowledge Collaboration in Online Communities," *Organization science* (22:5), pp. 1224-1239.

Faraj, S., and Johnson, S. L. 2011. "Network Exchange Patterns in Online Communities," *Organization Science* (22:6), pp. 1464-1480.

Farnham, S., Chesley, H. R., McGhee, D. E., and Kawal, R. 2000. "Structured Online Interactions: Improving the Decision-Making of Small Discussion Groups," in *Proceedings of the 2000 Conference on Computer-Supported Cooperative Work,* Philadelphia.

Farrell, J., and Saloner, G. 1985. "Standardization, Compatibility, and Innovation," *The RAND Journal of Economics* (16:1), pp. 70-83.

Figallo, C. 1998. *Hosting Web Communities: Building Relationships, Increasing Customer Loyalty, and Maintaining a Competitive Edge*, New York: John Wiley & Sons, Inc.

Fulk, J. 1993. "Social Construction of Communication Technology," *Academy of Management Journal* (36:5), pp. 921-950.

Galegher, J., Sproull, L., and Kiesler, S. 1998. "Legitimacy, Authority, and Community in Electronic Support Groups," *Written Communication* (15:4), pp. 493-530.

Hagel, J., and Armstrong, A. G. 1997. *Net Gain: Expanding Markets through Virtual Communities*, Boston: Harvard Business School Press.

Hall, H., and Graha, D. 2004. "Creation and Recreation: Motivating Collaboration to Generate Knowledge Capital in Online Communities," *International Journal of Information Management* (24:3), pp. 235-246.

Hamming, R. W. 1997. *The Art of Doing Science and Engineering: Learning to Learn*, Amsterdam: Gordon and Breach Science Publishers.

Hannan, M. T., and Freeman, J. 1984. "Structural Inertia and Organizational Change," *American Sociological Review* (49:2), pp. 149-164.

Herriott, S. R., Levinthal, D., and March, J. G. 1985. "Learning from Experience in Organizations," *American Economic Review* (75), pp. 298-302.

Hummel, J., and Lechner, U. 2002. "Social Profiles of Virtual Communities," in *Proceedings of the 35th Hawaii International Conference on System Sciences*, Los Alamitos, CA: IEEE Computer Society Press, pp. 2245-2254.

Hunter, J. E., Danes, J. E., and Cohen, S. H. 1984. *Mathematical Models of Attitude Change* (Volume 1), New York: Academic Press.

Jina, X. -L., Lee, M. K. O., and Cheung, C. M. K. 2010. "Predicting Continuance in Online Communities: Model Development and Empirical Test," *Behaviour & Information Technology* (29:4), pp. 383-394.

Johnston, B. "Online Community Metrics Best Practices Survey–March 2007," Forum One Communications, San Fransico.

Jones, Q., Moldovan, M., Raban, D. R., and Butler, B. 2008. "Empirical Evidence of Information Overload Constraining Chat Channel Community Interactions," in Proceedings of the 2008 Conference on *Computer-Supported Cooperative Work,* San Diego, CA, pp. 323-332.

Jones, Q., Ravid, G., and Rafaeli, S. 2004. "Information Overload and the Message Dynamics of Online Interaction Spaces: A Theoretical Model and Empirical Exploration," *Information Systems Research* (15:2), pp. 194-210.

Joyce, E., and Kraut, R. E. 2006. "Predicting Continued Participation in Newsgroups," *Journal of Computer Mediated Communication* (11:3), pp. 723-747.

Judge, T. A., and Bretz, R. D. 1992. "Effects of Work Values on Job Choice Decisions.," *Journal of Applied Psychology* (77), pp. 261-271.

Kane, G. C., and Alavi, M. 2007. "Information Technology and Organizational Learning: An Investigation of Exploration and Exploitation Processes," *Organization Science* (18:5), pp. 796-812.

Kane, G. C., Majchrzak, A., Johnson, J., and Chenisern, L. 2009. "A Longitudinal Model of Perspective Making and Perspective Taking Within Fluid Online Collectives," in *Proceedings of the 30th International Conference on Information Systems*, Phoenix, AZ.

Katz, M. L., and Shapiro, C. 1985. "Network Externalities, Competition, and Compatibility," *The American Economic Review* (75:3), pp. 424-440.

Kim, A. J. 2000. *Community Building on the Web*, Berkeley, CA: Peachpit Press.

Kollmann, T., and Krell, P. 2011. "Innovative Electronic Business: Current Trends and Future Potentials," *International Journal of E-Entrepreneurship and Innovation* (2:1), pp. 16-25.

Kraut, R. E., and Resnick, P. 2011. *Evidence-Based Social Design: Mining the Social Sciences to Build Successful Online Communities*, Cambridge, MA: MIT Press.

Kreps, D. M. 1990. "Corporate Culture and Economic Theory," in *Perspectives on Positive Political Economy*, J. Alt and K. A. Shepsle (eds.), Cambridge, UK: Cambridge University Press, pp. 90-143.

Kuk, G. 2006. "Strategic Interaction and Knowledge Sharing in the Kde Developer Mailing List," *Management Science* (57:7), pp. 1031-1042.

Lakhani, K. R., and von Hippel, E. 2003. "How Open Source Software Works: 'Free' User-to-User Assistance," *Research Policy* (32:6), pp. 923-943.

Lampel, J., and Bhalla, A. 2007. "The Role of Status Seeking in Online Communities: Giving the Gift of Experience," *Journal of Computer-Mediated Communication* (12:2), pp. 100-121.

Lant, T. K., and Mezias, S. 1990. "Managing Discontinuous Change: A Simulation Study of Organizational Learning and Entrepreneurship," *Strategic Management Journal* (11:5), pp. 147-179.

Lave, C., and March, J. G. 1975. *An Introduction to Models in the Social Sciences* New York, NY: Harper and Row.

Law, A. M., and Kelton, W. D. 1991. *Simulation Modeling and Analysis* (2nd ed.), New York: McGraw-Hill.

Law, J., and Singleton, V. 2005. "Object Lessons," *Organization Science* (12:3), pp. 331-355.

Lazar, J., and Preece, J. 2003. "Social Considerations in Online Communities: Usability, Sociability, and Success Factors," in *Cognition in a Digital World*, H. Van Oostendorp (ed.), Mahwah, NJ: Lawrence Erlbaum Associates, pp. 112-132.

Levine, J. M., and Moreland, R. L. 1999. "Knowledge Transmission in Work Groups: Helping Newcomers to Succeed," in *Shared Cognition in Organizations: The Management of Knowledge*, L. Thompson, J. Levine, and D. Messick (eds.), Mahwah, N J: Lawrence Erlbaum Associates, pp. 267-296.

Ludford, P. J., Cosley, D., Frankowski, D., and Terveen, L. 2004. "Think Different: Increasing Online Community Participation Using Uniqueness and Group Dissimilarity," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vienna, Austria.

Ma, M., and Agarwal, R. 2007. "Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities," *Information Systems Research* (18:1), pp. 42-67.

Majchrzak, A., Wagner, C., and Yates, D. 2013. "The Impact of Shaping on Knowledge Reuse for Organizational Improvement with Wikis," *MIS Quarterly* (37:2), pp. 455-469.

March, J. G. 1991. "Exploration and Exploitation in Organizational Learning," *Organization Science* (2:1), pp. 71-87.

Markus, M. L. 1987. "Toward A "Critical Mass" Theory of Interactive Media," *Communication Research* (14:5), pp. 491-511.

McAfee, A. 2006. "Mastering the Three Worlds of Information Technology," *Harvard Business Review* (84:11), pp. 141-149.

McPherson, J. M. 1983. "The Size of Voluntary Organizations," *Social Forces* (61:4), pp. 1045-1064.

Millen, D. R., Fontaine, M. A., and Muller, M. J. 2002. "Understanding the Benefits and Costs of Communities," *Communications of the ACM* (45:4), pp. 64-68.

Morecroft, J. D. 1985. "Rationality in the Analysis of Behavioral Simulation Models," *Management Science* (31:7), pp. 900-916.

Moreland, R. L., and Levine, J. M. 1982. "Socialization in Small Groups: Temporal Changes in Individual-Group Interactions," *Social Psychology* (15), pp. 137-192.

Nikolai, C., and Madey, G. 2009. "Tools of the Trade: A Survey of Various Agent Based Modeling Platforms," *Journal of Artificial Societies and Social Simulation* (12:2).

Nov, O., and Arazy, O. 2013. "Personality-Targeted Design: Theory, Experimental Procedure, and Preliminary Results," in *Proceedings of the 2013 Conference on Computer-Supported Cooperative Work*, San Antonio, Texas, Febuary 23-27.

Oh, W., and Jeon, S. 2007. "Membership Herding and Network Stability in the Open Source Community: The Ising Perspective," *Management Science* (53:7), pp. 1086-1101.

Orlikowski, W., and Iacono, S. 2001. "Desperately Seeking the 'IT' in IT Research," *Information Systems Research* (7:4), pp. 400-408.

Peddibhotla, N. B., and Subramani, M. R. 2007. "Contributing to Public Document Repositories: A Critical Mass Theory Perspective," *Organization Studies* (28:3), pp. 327-349.

Phang, C. W., Kankanhalli, A., and Sabherwal, R. 2009. "Usability and Sociability in Electronic Communities: A Comparative Study of Knowledge Seekers and Contributors," *Journal of the Association for Information Systems* (10:10), pp. 721-747.

Posner, B. Z. 1992. "Person–Organization Values Congruence: No Support for Individual Differences as a Moderating Influence," *Human Relations* (45:4), pp. 351-361.

Preece, J. 1999. "Empathy Online," *Virtual Reality* (4:1), pp. 74-84.

Preece, J. 2000. *Online Communities: Supporting Sociability, Designing Usability*, Chichester, UK: John Wiley and Sons.

Preece, J. 2001. "Sociability and Usability in Online Communities: Determining and Measuring Success," *Behaviour & Information Technology* (20:5), pp. 347-356.

Ransbotham, S., and Kane, G. C. 2011. "Membership Turnover and Collaboration Success in Online Communities: Explaining Rises and Falls from Grace in Wikipedia," *MIS Quarterly* (35:3), pp. 613-627.

Rashid, A. M., Ling, K., Tassone, R. D., Resnick, P., Kraut, R., and Riedl, J. 2006. "Motivating Participation by Displaying the Value of Contribution," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York: ACM Press, pp. 955-958.

Ren, Y., Harper, F. M., Drenner, S., Terveen, L., Kiesler, S., Riedl, J., and Kraut, R. E. 2012. "Building Member Attachment in

Online Communities: Applying Theories of Group Identity and Interpersonal Bonds," *MIS Quarterly* (36:3), pp. 841-864.

Resnick, P., Konstan, J., Chen, Y., and Kraut, R. 2011. "Starting New Online Communities," in *Evidence-Based Social Design: Mining the Social Sciences to Build Online Communities,* R. E. Kraut and P. Resnick (eds.), Cambridge, MA: MIT Press.

Rheingold, H. 1993. *The Virtual Community: Homesteading on the Electronic Frontier*, Reading, MA: Addison-Wesley.

Ridings, C. M., and Gefen, D. 2004. "Virtual Community Attraction: Why People Hang Out Online," *Journal of Computer-Mediated Communication* (10:1).

Ridings, C., and Wasko, M. 2010. "Online Discussion Group Sustainability: Investigating the Interplay between Structural Dynamics and Social Dynamics over Time," *Journal of the Association of Information Systems* (11:2), pp. 95-121.

Robey, D., and Boudreau, M. C. 1999. "Accounting for the Contradictory Organizational Consequences of Information Technology: Theoretical Directions and Methodological Implications," *Information Systems Research* (10:2), pp. 167-185.

Ryan, A. M., Sacco, J. M., McFarland, L. A., and Kriska, S. D. 2000. "Applicant Self-Selection: Correlates of Withdrawal from a Multiple Hurdle Process," *Journal of Applied Psychology* (85:2), pp. 163-179.

Rynes, S. L., Bretz, R. D., and Gerhart, B. 1991. "The Importance of Recruitment in Job Choice: A Different Way of Looking," *Personnel Psychology* (44:3), pp. 487-521.

Sabherwal, R., and Robey, D. 1993. "An Empirical Taxonomy of Implementation Processes Based on Sequences of Events in Information Systems Development," *Organization Science* (4:4), pp. 548-576.

Schau, H. J., Muniz Jr., A. M., and Arnould, J. E. J. 2009. "How Brand Website Practices Create Value," *Journal of Marketing* (73:3), pp. 30-51.

Schneider, B. 1987. "The People Make the Place," *Personnel Psychology* (40:3), pp. 437-453.

Schneider, B., Goldstein, H. W., and Smith, D. B. 1995. "The ASA Framework: An Update," *Personnel Psychology* (48:4), pp. 747-773.

Scott, W. R., and Davis, G. F. 2007. *Organizations and Organizing: Rational, Natural and Open Systems Perspectives* Upper Saddle River, NJ: Prentice Hall.

Sharma, N., Butler, B. S., Irwin, J., and Spallek, H. 2011. "Emphasizing Social Features in Information Portals: Effects on New Member Engagement," *Journal of the Association of Information Sciences and Technology* (62:11), pp. 2106-2120.

Shen, K. N., and Khalifa, M. 2009. "Design for Social Presence in Online Communities: A Multidimensional Approach," *AIS Transactions on Human-Computer Interaction* (1:2), pp. 33-54.

Spagnoletti, P., and Resca, A. 2012. "A Design Theory for IT Supporting Online Communities," in *Proceedings of the 45th Hawaii International Conference on System Sciences*, Los Alamitos, CA: IEEE Computer Society Press, pp. 4082-4091.

Sproull, L., and Kiesler, S. 1991. *Connections: New Ways of Working in the Networked Organization*, Boston: MIT Press.

Thompson, M. 2005. "Structural and Epistemic Parameters in Communities of Practice," *Organization Science* (16:2), pp. 151-164.

Turban, D. B., and Keon, T. L. 1993. "Organizational Attractiveness: An Interactionist Perspective," *Joumal of Applied Psychology* (78), pp. 184-193.

Turner, J. H. 1988. *A Theory of Social Interaction*, Stanford, CA: Stanford University Press.

Turner, T. C., Smith, M. A., Fisher, D., and Welser, H. T. 2005. "Picturing Usenet: Mapping Computer-Mediated Collective Action," *Journal of Computer-Mediated Communication* (10:4).

Van Voorhis, C. R. W., and Morgan, B. L. 2007. "Understanding Power and Rules of Thumb for Determining Sample Sizes," *Tutorials in Quantitative Methods for Psychology* (3:2), pp. 43-50.

Venable, J. 2006. "The Role of Theory and Theorizing in Design Science Research," in *Proceedings of the 1st International Conference on Design Science in Information Systems and Technology*, Claremont, CA, pp. 1-18.

Viegas, F. B., Wattenberg, M., Kriss, J., and van Ham, F. 2007. "Talk Before You Type: Coordination in Wikipedia Export," in *Proceedings of the 40th Hawaii International Conference on System Sciences*, Los Almitos, CA: IEEE Computer Society Press.

Vroom, V. R. 1966. "Organizational Choice: A Study of Pre- and Post-Decision Processes," *Organizational Behavior and Human Performance* (1:2), pp. 212-226.

Walther, J. B. 1996. "Computer-Mediated Communication," *Communication Research* (23:1), pp. 3-43.

Wanous, J. P., Poland, T. D., Premack, S. L., and Davis, K. S. 1992. "The Effects of Met Expectations on Newcomer Attitudes and Behaviors: A Review and Meta-Analysis," *Journal of Applied Psychology* (77), pp. 288-297.

Wasko, M. M., and Faraj, S. 2000. "It Is What One Does: Why People Participate and Help Others in Electronic Communities of Practice," *The Journal of Strategic Information Systems* (9:2-3), pp. 155-173.

Wasko, M. M., and Faraj, S. 2005. "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice," *MIS Quarterly* (29:1), pp. 35-57.

Wasserman, S., and Faust, K. 1994. *Social Network Analysis: Methods and Application*, Cambridge, UK: Cambridge University Press.

Welser, H. T., Gleave, E., Fisher, D., and Smith, M. 2007. "Visualizing the Signatures of Social Roles in Online Discussion Groups," *Journal of Structured Systems* (8).

Worthen, B. 2008. "Why Most Online Communities Fail," *Wall Street Journal*, July 16.

Zammuto, R. F., Griffith, T. L., Majchrzak, A., Dougherty, D. J., and Faraj, S. 2007. "Information Technology and the Changing Fabric of Organization," *Organization Science* (18:5), pp. 749-762.

Zhang, X., and Zhu, F. 2011. "Group Size and Incentives to Contribute," *The American Economic Review* (101), pp. 1601-1615.

Zott, C. 2003. "Dynamic Capabilities and the Emergence of Intra-Industry Differential Firm Performance: Insights from a Simulation Study," *Strategic Management Journal* (24:2), pp. 97-125.

About the Authors

Brian S. Butler is an associate professor at the College of Information Studies and the Robert H. Smith School of Business at the University of Maryland. His research interests include IT impacts on local food systems and other geographically embedded markets, business continuity management, mindfulness and IT, the dynamics of online communities and other technology-supported groups, Wikipedia policies and rules, social networking systems, the politics of technology implementation in organizations, and the impact of electronic commerce on interorganizational relationships. His research has been published in venues such as *MIS Quarterly*, *Information Systems Research*, *Organization Science*, *Journal of Medical Internet Research, Information Technology & People*, and *Journal of Management Information Systems*.

Patrick J. Bateman is an associate professor at the Williamson College of Business Administration, Youngstown State University. His research focuses on understanding the dynamic interactions between people and technologies, in particular individuals' use of interactive and socio-technical systems, including social networking, online communities, virtual worlds, and other forms of emerging computer-mediated communication. His research has been published in range of leading journals, including *MIS Quarterly*, *Information Systems Research*, *Journal of Computer-Mediated Communication,* and *Information Technology & People*.

Peter H. Gray is an associate professor at the McIntire School of Commerce, University of Virginia. His research focuses on the collaborative impacts of social technologies, organizational networks, virtual teams, online communities, and knowledge management systems. His research has been published in a range of leading journals, including *MIS Quarterly*, *Information Systems Research*, *Management Science*, *Sloan Management Review*, *Organizational Research Methods*, *California Management Review*, *Journal of Management Information Systems*, *Journal of Strategic Information Systems*, and *Information Technology & People*.

E. Ilana Diamant is a post-doctoral research scientist at the University of Washington's department of Human Centered Design & Engineering. Her research investigates collaboration issues in the development and use of large-scale systems, such as information infrastructures for e-science, and Open Source Software.

Appendix A

Table A1. Community Parameters and Calibrated Values

Community Parameters	*Participation Cost (PC)*	Within the community platform, the ratio of the per-message processing cost over the average benefit provided by a message deemed interesting (Range: [0,1])	0.33
	Topic Consistency Cues (TCC)	The degree to which community technology features signal that current message content is indicative of future content (Range: [0,1])	0.02
	Attraction Rate (A)	The number of interested individuals who join the community in each time period	Randomly selected from a gamma distribution with parameters a = 0.2624 and b = 1.1264
	Initial Fit Expectation Distribution	Expectations formed prior to joining about the degree of fit between an individual's interests and the communication activity within the community (Range: 0 to 1).	Uniformly distributed between 0.75 and 1
	Initial Volume Expectation Distribution	Expectations formed prior to joining about the number of messages present in the community each period (Range: $V_0 \geq 0$)	1 for all individuals
	Message Contribution probability Distribution	The likelihood of the individual contributing a message to the community (Range: 0 to 1)	Ratio of participants (i.e., individuals for whom $p_i > 0$) to members is chosen from an gamma distribution with parameters of a = 0.3024 and b = 0.5268; All individuals labeled as participants have the same participation probability, a value that is drawn from a log-normal distribution [LN(-4.10,0.55)
	Interest Distribution	The set of topics that an individual is interested in (A range of values between 0 and 1)	Individuals' interest range length is chosen from a uniform distribution between 0 and the community's maximum interest range, a value that is selected from a uniform distribution between 0.25 and 0.75

Table A2. Descriptive Statistics

	N	Min	Max	Mean	Std. Dev
Percentage Membership Loss	192	0	0.98	.131	.168
Group size (people)	192	3.0	225	163.59	282.06
Communication volume (messages/day)	192	0	29.13	1.11	3.05
Number of people entering per day	192	0	4.07	.31	.646