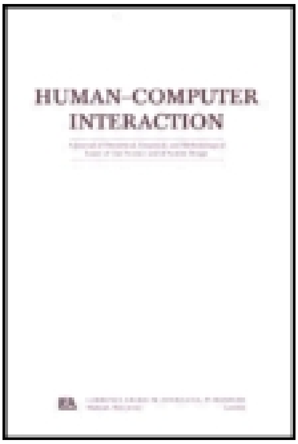


Human-Computer Interaction
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hhci20

Agent-Based Modeling to Inform Online Community Design: Impact of Topical Breadth, Message Volume, and Discussion Moderation on Member Commitment and Contribution

Yuqing Ren[a] & Robert E. Kraut[b]
[a] University of Minnesota
[b] Carnegie Mellon University
Accepted author version posted online: 15 Oct 2013.Published online: 29 Apr 2014.

HUMAN–COMPUTER INTERACTION, 2014, Volume 29, pp. 351–389

ISSN: 0737-0024 print / 1532-7051 online
DOI: 10.1080/07370024.2013.828565

Taylor & Francis
Taylor & Francis Group

Agent-Based Modeling to Inform Online Community Design: Impact of Topical Breadth, Message Volume, and Discussion Moderation on Member Commitment and Contribution

Yuqing Ren[1] **and Robert E. Kraut**[2]
[1]*University of Minnesota*
[2]*Carnegie Mellon University*

The design of complex social systems, such as online communities, requires the consideration of many parameters, a practice at odds with social science research that focuses on the effects of a small set of variables. In this article, we show how synthesizing insights from multiple, narrowly focused social science theories in an agent-based model helps us understand factors that lead to the success of online communities. The agent-based model combines insights from theories related to collective effort, information overload, social identity, and interpersonal attraction to predict motivations for online community participation. We conducted virtual experiments to develop hypotheses around three design decisions about how to orchestrate an online community—topical breadth, message volume, and discussion moderation—and the trade-offs involved in making these decisions. The simulation experiments suggest that broad topics and high message volume can lead to higher member commitment. Personalized moderation outperforms other types of moderation in increasing members' commitment and contribution, especially in topically broad communities and those with high message volume. In comparison,

Yuqing Ren (chingren@umn.edu, www.tc.umn.edu/∼chingren) is a social scientist with an interest in online community design, social media, distributed collaboration, knowledge management, and computational modeling of social and organizational systems; she is an Assistant Professor of Information and Decision Sciences at the Carlson School of Management at the University of Minnesota. **Robert E. Kraut** (robert.kraut@cmu.edu, kraut.hciresearch.org) is a social psychologist with an interest in the design and social impact of computing, analysis, and design of online communities, everyday use of the Internet, technology and conversation, collaboration in small work groups, and computing in organizations; he is the Herbert A. Simon Professor in the School of Computer Science and the Tepper School of Business at Carnegie Mellon University.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hhci.

CONTENTS

community-level moderation increases commitment but not contribution, and only in topically narrow communities. These simulation results suggest a critical trade-off between informational and relational benefits. This research illustrates that there are many interactions among the design decisions that are important to consider; the particulars of the community's goals often determine the effectiveness of some decisions. It also demonstrates the value of agent-based modeling in synthesizing simple social science theories to describe and prescribe behaviors in a complex system, generating novel insights that inform the design of online communities.

1. INTRODUCTION

The Internet provides a popular platform for social interactions. According to the Pew Internet and American Life Project, even in 2001, 84% of adult Internet users in the United States, or about 90 million people, participated in some type of online groups to share information; exchange social support; and discuss hobbies, politics, sports, and other topics of interest. In 2012, 66% of adult Internet users in the United States use a social networking site like Facebook or LinkedIn, and 48% use it on a daily basis. Following Preece (2000), we define an online community as an Internet-connected collective of people interacting over time around a shared purpose, interest, or need. Although some online communities are highly successful, many others fail. For example, 20% of newsgroups studied were entirely empty, and 42% had fewer than 100 messages in a 10-week period (Smith, 1999). Similarly, half of Internet Relay Chat channels that appeared in July 2005 lasted no more than 1 day, and 97% died within 6 months (Raban, Moldovan, & Jones, 2010). In the business world, failure rates are also alarmingly high. A Deloitte survey of more than 100 businesses attempting to build online communities—some spending more than $1 million in the effort—found that most efforts failed to attract a critical mass of users. About 35% of the communities had fewer than 100 users and 75% had fewer than 1,000 users (Worthen, 2008).

An important reason behind these failures is the lack of evidence-based guidance for building and managing online communities. Designers and managers must make numerous decisions about features, structures, and policies to build a successful community. Even experienced designers can get overwhelmed by the trade-offs involved in making these decisions and fail to anticipate how users will respond. For instance, the moderation and removal of off-topic messages may encourage some participants to visit a community more frequently but may discourage others from participating or cause them to leave the community.

These effects of off-topic moderation originate from two social science theories: information overload theory (Jones, Ravid, & Rafaeli, 2004) and interpersonal bonds theory (Collins & Miller, 1994). The former argues that off-topic messages do not provide informational value and sorting through volumes of irrelevant content causes informational overload. The latter argues that off-topic messages provide a good opportunity for members to share personal stories and engage in self-disclosure, leading to interpersonal relationships. Removing off-topic messages therefore reduces relationship development.

Although theories from social psychology, organizational behavior, sociology, and economics have been applied to describe behaviors in online communities, few have been applied prescriptively to offer recommendations for building successful communities (see Kraut & Resnick, 2012; Ling et al., 2005, for exceptions). An important reason, we suspect, may be that the logic of design, which involves trade-offs among tens of parameters that could influence participant behavior, is at odds with the logic of social science research, which examines the influence of a small set of variables while holding everything else equal. Applying social science theory to design requires a way to synthesize insights from multiple theories and identify the pathways through which particular design choices affect the outcomes that designers aim to achieve.

In this article, we advocate a new approach to both theory development and design of online communities by synthesizing propositions from constituent social science theories into an agent-based model. Our approach is similar to what Allen Newell (1973) advocated in his famous "20 questions" paper. In that paper, he argued that the field of cognitive psychology, in those days, was making slow progress because it focused on discovering and mapping discrete phenomena. Instead, Newell advocated using simulation models to represent and unify theories. This approach has led to more comprehensive and unified theory in cognitive psychology (Anderson, 1996). We extend this approach to social sciences and demonstrate its usefulness by developing and applying the model to understand design decisions in conversation-based online communities.[1] A common challenge that online community designers face is balancing the quantity and quality of information flow by, for example, carefully defining the community's niche. Online communities with too narrow a niche (e.g., web forum for the movie *Mulan*) may not attract a "critical mass" of members needed to be successful. Conversely, online communities with too broad a niche (e.g., IMDb discussion forums of anything about movies) can easily overwhelm members with too many messages, most of which are not of interest to any particular one of them. Consequently, members may have trouble finding worthwhile messages buried in the deluge of messages (Lampe, Johnston, & Resnick, 2007), and information overload occurs. Community members who experience information overload are more likely to leave the community if their expected costs of processing the information exceed expected benefits (Jones et al., 2004).

In this article, we developed an agent-based model drawing insights from social science theories and empirical data from UseNet groups. We used the model to examine three actions online community designers can take to manage the challenge of balancing the quantity and quality of communication within an online community: (a) defining the appropriate topical breadth of the community, (b) changing the cost of contribution to manage message volume, and (c) introducing technical and social means to filter and moderate the discussion. We simulated three levels of topical breadth and message volume, and three ways of moderation—no moderation, community-level moderation, and personalized moderation—and examined how they affect member commitment and community activity. Simulation results challenge previously established guidelines about keeping community topics narrow and focused. Instead, they suggest that online communities thrive with high volumes of message traffic covering a broad range of topics if they also use personalized

[1]An online community can be created on various technological platforms (e.g., listservers, *Usenet* news, chats, bulletin boards, web forums, and social networking sites) around various purposes (e.g., interest, health support, technical support, education, e-commerce; Preece 2000). In this article, we focus on conversation-based interest communities such as newsgroups or web forums created to host online discussion of shared interests.

moderation as a way to balance diverse user interests. The results also reveal a critical trade-off between designing for informational versus relational benefits.

The rest of the article is organized as follows. In Section 2, we introduce agent-based modeling as a tool to combine insights from multiple theories for understanding trade-offs in online community design. In Section 3, we present some challenges of managing online conversations and three design decisions to address them. We then describe the conceptual framework for the model, its implementation and validation, and simulation experiments and results. Finally, we discuss how extensions of the model can serve as a test bed to inform online community design.

2. THEORY INTEGRATION IN AGENT-BASED MODELS

Scholars have used many social science theories to understand what makes online communities successful. Ren, Kraut, and Kiesler (2007) and Sassenberg (2002) applied theories of group identity and interpersonal bonds to examine the development of members' commitment to online communities. Jones et al. (2004) and Butler (2001) used theories of information overload to examine the effects of community dynamics on member behaviors. Kollock (1999) and Ling et al. (2005) applied public goods economics and theories of social loafing to analyze problems of undercontribution.

Rarely have social science theories been used prescriptively as the basis for designing online communities. A major reason is that the logic of design, which attempts to manage trade-offs among many parameters that can influence a community's success, is at odds with the logic of social science research, which attempts to examine the influence of a small set of variables while holding everything else equal. This ceteris paribus paradigm for developing and testing social science hypotheses produces theories that are often too simple for the purpose of social engineering. Social science studies, even if they examine many variables simultaneously, rarely examine higher order interactions. In contrast, social engineering requires theory that describes the behavior of a large set of factors varying simultaneously and their interactions over a long time.

In this article, we use an agent-based model to integrate what Davis, Eisenhardt, and Bingham (2007) termed "simple theories" to understand factors that affect online community success. Agent-based models capture the behaviors of complex adaptive systems by modeling the behavior of the individuals who comprise them (North & Macal, 2007). The emergent properties of a complex social system (e.g., a beehive, a financial market, or an online community) are examined by simulating the behaviors of the agents that comprise the collective (e.g., the bees, traders, or members). Compared with conventional methods, agent-based modeling is especially suitable for bottom-up theorizing (Kozlowski & Klein, 2000), and is useful for understanding how individual behaviors interact over time and lead to emergent system-level patterns.

The model simulates people's motivation to participate in and contribute to an online community. Its core assumption is that people participate and contribute to the extent they believe their efforts will lead to outcomes they value. We base the model on the expectancy-value theory of motivation (Vroom, Porter, & Lawler, 2005) and one of its extensions, the collective effort model (Karau & Williams, 1993). Expectancy-value theories, however, are silent about the different benefits or valued outcomes members derive from participating in a group and need to be complemented by insights from other bodies of literature.

Prior research has identified six benefits that have been consistently found to motivate participation in online communities: (a) information; (b) fulfillment of altruistic or expressive needs produced by helping others; (c) identification with the group; (d) friendships or relationships formed with group members; (e) fun, entertainment, and other forms of intrinsic motivation; and (f) reputation and other forms of extrinsic motivation (e.g., Ren et al., 2007; Ridings & Gefen, 2004; Roberts, Hann, & Slaughter, 2006; Wasko & Faraj, 2005).

Each of these motivators can be explained by separate social science theories about how they operate. For example, theories of group identity and interpersonal bonds propose that members will be committed to a group and will contribute to it if they feel psychologically attached to the group as a whole (e.g., Hogg, 1996) or its members (Lott & Lott, 1965). Resource-based theory (Butler, 2001) proposes that people participate in groups to get access to the information that other group members provide, whereas information overload theory (Jones et al., 2004) proposes that people's information-processing capacity is limited and too much information or noise is aversive (Rogers & Agarwala-Rogers, 1975). Figure 1 includes an overview of the theories, their key assumptions, and examples of their conflicting predictions. We constructed the theoretical framework to be representative rather than exhaustive; when multiple theories apply, we selected one to represent each motivation.

Figure 1 illustrates the challenge of online community design and the value of using agent-based modeling to combine multiple theories to inform design. First, multiple causes converge to determine people's overall motivation to participate in an online group, and separate social science theories have been built to account for each cause. For example, information overload theory focuses on how informational benefits affect motivation, whereas group identity theory focuses on psychological attachment to the community. Therefore, multiple theories are needed to model motivation in online communities and to predict the effects of various interventions.

Second, because theories focus on different underlying processes, they often predict that a single design choice will have different effects on motivation to contribute to an online community. For example, theories of interpersonal attraction propose that people like others who are similar to themselves (Byrne, 1961). Therefore, increasing group homogeneity may cause members to contribute *more* because people like and are willing to expend effort to help similar others. On the other hand, the collective effort model proposes that people contribute more when they believe their contributions are unique and indispensable for group success. As a result, increasing group homogeneity may cause members to contribute *less* because members feel their efforts are redundant. Another example is the effect of group size. When examined through Butler's (2001) resource-based theory of online social groups,

FIGURE 1. Overview of constituent social theories, key assumptions, and conflicting predictions.

Motivations	Social Theories	Key Assumptions	Conflicting Predictions From Theories
Overall motivation derives from personal benefits of participating in a group	Expectancy theory (Vroom et al., 2005) Collective effort model (Karau & Williams, 1993)	Individual motivation depends upon expectancy (that effort leads to performance), instrumentality (that performance leads to outcomes), and valence (that the outcomes are desirable) • Members contribute more when group valence is high • Members contribute more when task valence is high • Members contribute more when they think their effort is instrumental to group outcomes	• Effects of ***Group Size*** ○ Collective effort model predicts large size decreases motivation due to free riding and dilution of responsibility ○ Resource-based theory predicts large size increases motivation by increasing information benefit ○ Interpersonal bonds theory predicts large size decreases motivation by reducing the chance to know other members • Effects of ***Group Similarity*** ○ Collective effort model predicts similarity decreases motivation because of perceived redundancy in contribution (lack of uniqueness of contribution) ○ Information overload theory predicts similarity increases motivation because of access to information that matches one's interest ○ Interpersonal bonds theory predicts similarity increases motivation because of greater chance of liking one another • Effects of ***Personal Disclosure*** ○ Information overload theory predicts personal disclosure decreases motivation because it increases noise/signal ratio ○ Interpersonal bonds theory predicts personal disclosure increases motivation by strengthening relationships among members
Benefit from accessing information	Resource-based theory (Butler, 2001) Information overload theory (Jones et al., 2004)	Human beings' information-processing capacity is limited and too many messages increase processing cost • Only messages that match members' interests provide information benefit • Benefit from accessing information is a marginally decreasing function of quantity	
Benefit from helping and sharing information	Collective effort model (Karau & Williams, 1993)	Individuals make contributions to help others or a group • Members contribute more in small groups • Members contribute more when they think their contribution is unique or others are undercontributing	
Social benefit from attachment to a group as a whole	Group identity (Hogg & Terry, 2000)	Members are likely to contribute to groups if they feel psychologically attached to the group as a whole • Attachment is stronger with social categorization, group homogeneity, and intergroup comparison	
Social benefit from attachment to group members	Interpersonal bonds (Collins & Miller, 1994; Festinger et al., 1950)	Members are likely to contribute to groups if they feel psychologically attached to its members • Stronger attachment with personal disclosure, interpersonal similarity, and repeated interaction	
Enjoyment	Intrinsic motivation (Deci & Ryan, 2000)	Members perform an activity for its own sake because it is interesting and they get pleasure from doing it • Enjoyment varies across activities and individuals	
Reputation	Extrinsic motivation (Deci & Ryan, 2000)	Members perform an activity due to external incentives such as rewards, coercion, or threat of punishment • Greater motivation from high ranking or reputation	

large group size is a measure of resource availability and thus provides the benefits of increased information access. When examined through the collective effort model (Karau & Williams, 1993), members of large groups tend to contribute less time and resources because of dilution of responsibility. Combining these paths in an agent-based model, and allowing them to interact over time in a nonlinear fashion can lead to a deeper understanding of how various design decisions affect member motivation and contribution. In this article, we illustrate this use of agent-based modeling by examining three critical decisions in online community design.

3. CHALLENGES OF ONLINE COMMUNITY DESIGN

At the core of most online communities are members who converse to ask and answer questions, exchange opinions and social support, develop friendships, or just pass the time. Without conversation, these communities could not function. Even in online games like *World of Warcraft* or production-oriented communities like *Wikipedia*, members depend upon conversation to coordinate activity or build relationships. Although communication is central to most online communities, too much communication or the wrong kind can threaten them. Information overload fueled by high message volume and irrelevant topics can drive people away from online communities (Butler, 2001; Jones et al., 2004). According to Jones et al. (2004), information overload occurs when an individual (or system) cannot process and use all communication input or when the effort required to process the information exceeds the amount members are prepared to invest. In online conversations, information overload can be caused by *conversational overload*—when too many messages are posted—or *information entropy*—when messages are not sufficiently organized by topics or as part of a conversation. That is, in many online communities, information overload increases with message volume or the percentage of messages that do not match a member's interest (Lackaff, 2005).

In many online communities, personal, off-topic conversations that are irrelevant to the nominal topic represent "noise" for a typical community member. For example, these might be messages in a depression discussion group having little to do with depression (http://discussions.seniornet.org; Wright, 2000), or messages in investment discussion groups having little to do with finance (Gu, Konana, Rajagopalan, & Chen, 2007). These off-topic messages are an irritation for people who are interested only in the nominal topic of the community and can drive them away. High message volume is also problematic in communities that encourage conversation across a wide range of topics. As Butler (1999) noted, in topically diverse communities, messages that are interesting to some community members are likely to be uninteresting to others.

3.1. Design Decisions to Manage Information Overload

Individual users can adopt a range of actions to reduce the impact of information overload, such as increasing their effort, focusing on a narrow set of topics while ignoring others, or ending active participation (Jones, Ravid, & Rafaeli, 2002). At the same time, community designers have several options to reduce information overload and help readers focus their attention on messages likely to interest them (Resnick, Iacovou, Suchak, Bergstrom, & Riedl, 1994). Here we consider some straightforward choices that community designers can make to deal with information overload and their wider implications for the success of the community.

One option is to create specialized online groups that focus on a single or few topics. Some researchers believe highly specialized groups are the only reasonable way to cope with a plethora of online groups (e.g., Van Alstyne & Brynjolfsson, 1996). A second option is to limit message volume by controlling who can post a message or how many they can post. For example, the community can be public, allowing anyone to post, or private, with posting privileges limited to registered members. Communities can also impose explicit throttles on the number of posts a user can make per unit of time or raise the costs of posting. Researchers, for example, have proposed adding postage to messages to reduce volume and increase quality (Kraut, Sunder, Telang, & Morris, 2005).

A third option that designers can employ to reduce overload is to filter or moderate discussion based on content, allowing only messages relevant to the community. Moderation could occur at either the community level or the user level. A common practice is community-level moderation, where human moderators or software agents block or remove inappropriate or off-topic messages (Figallo, 1998; Lampe & Johnston, 2005). In this case, a message can be seen either by everyone visiting the site or by no one. Community-level moderation can be performed ex ante, by approving or rejecting messages before they can be posted, or ex post, by removing messages after they have been posted. The goal is to prevent off-topic or other inappropriate messages, such as spam, trolling messages, or antisocial flames. In contrast, with personalized moderation[2] different users view different subsets of messages matched to their interests. One classic example is *Usenet* killfile, which allows a user to ignore a set of messages based on simple criteria such as keywords or the poster's name (Lackaff, 2005). Collaborative filtering algorithms have been developed that recommend newsgroup articles based on users' previous ratings of other articles (Resnick et al., 1994). Software agents also can match messages against a static personal profile or one that is dynamically updated based on a user's behavior in the community (e.g., Harper et al., 2007). Social networking sites like Facebook and Google Plus use imputed tie strength as well as topics to determine which messages a user sees.

3.2. The Designer's Dilemma

From the perspective of community designers, choosing the best options for dealing with information overload can be a daunting task, because each option has

[2]We examine personalized moderation at the conceptual level. Its implementation is beyond the scope of this article.

implications for a variety of valued outcomes, such as attracting and retaining members, building relationships among members, or developing a valuable information repository. Consider the decision about defining the topical breadth of the community. Some researchers believe that a narrow focus is important for the success of an online community. By defining the community around a narrow focus, most participants will be interested in the majority of messages in the community. Maloney-Krichmar and Preece (2005) studied a support group for patients with anterior cruciate ligament knee injuries and attributed the vibrant nature of the community to its narrow focus. The narrow focus promotes the development of strong feelings of closeness due to the ease of finding shared interests (Schneider, Goldstein, & Smith, 2006; Walther, 1994). Some researchers suggest that a narrow topical focus accelerates the growth of a new online community because it lets newcomers know what to expect and recruits a homogeneous community of members (Wang & Kraut, 2012). Other researchers, however, have suggested issues with organizing a community around a narrow topical focus. By definition, a community organized around a single narrow topic, such as action movies, will appeal to fewer potential members than one organized around broader topics like all movies, or complementary topics like movies and TV shows. Therefore, a narrow topical focus, especially early in its history, may reduce the chance of the community reaching self-sustaining critical mass (Allen, 1988; Markus, 1987). Too narrow a focus may also limit opportunities for interactions or relationship development. For example, Wellman and Gulia (1999) observed that members of an online BMW group knew little about each other besides the model of car they drove and their repair expertise because the group had rules forbidding comments unrelated to BMWs.

Similar considerations come into play when community designers decide the style and amount of moderation to impose on conversation on the site. Even though community-level moderation is common, user-level moderation may be more effective in retaining members, especially in communities that are broadly defined. In a broadly defined community, some nominally on-topic messages will be of little interest to a large proportion of members. For example, in the movie discussion site JoBlo.com, a message evaluating a new action movie is likely to be of little interest to members who are only interested in romantic comedies. Conversely, nominally off-topic conversations, such as descriptions of high school romances consummated in movie theatres, may be of great interest to a small group of members. Under either scenario, community-level moderation leads to a suboptimal user experience. On the other hand, there are concerns that user-level moderation promotes individual convenience over community health, because members each see a subset of messages and do not have a common view of the community (Lackaff, 2005).

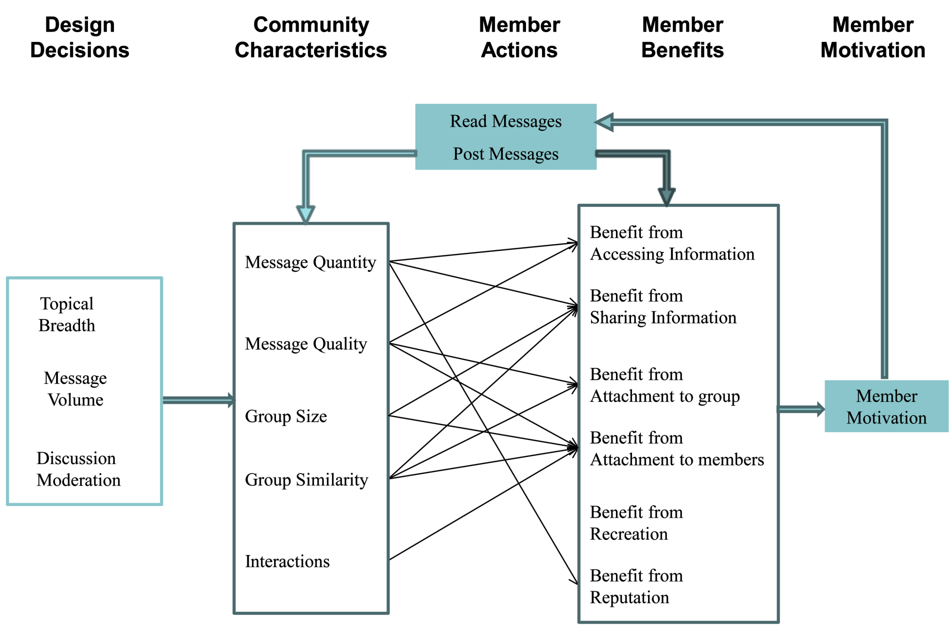
In this article, we examine these trade-offs involved in managing online conversations with three design decisions regarding topical breadth, message volume, and discussion moderation. By representing an online community as an agent-based model that synthesizes constituent social science theories, we aim to answer three questions: (a) How do topical breadth, message volume, and moderation style affect a community's viability and its members' commitment? (b) To what extent are the effects of moderation contingent upon community characteristics such as topical breadth and message volume? (c) How do the design decisions affect trade-offs among the various benefits members receive from participating in an online community?

4. CONCEPTUAL FRAMEWORK OF THE AGENT-BASED MODEL

Figure 2 depicts the conceptual framework of the agent-based model. The model combines insights from theories such as the collective effort model, information overload, and theories of group identity and interpersonal bonds to represent a composite theory of member motivation. Member actions like reading and posting messages are determined by benefits and costs associated with participation. Reading and posting behaviors change community dynamics such as the number and quality of messages, community size, and relationships among members; these, in turn, influence experienced benefits and motivation. Design interventions, such as the cost of posting messages, diversity of topics, and moderation also influence community characteristics.

Due to its complexity, we describe the model in three steps. We first describe how we calculate the various benefits the agent receives from participation. Then we describe the decision rules that determine whether an agent reads or posts a message. These decision rules are primarily based on insights from existing social

FIGURE 2. The conceptual framework of the agent-based model.



science theories and empirical analysis of 100 Usenet groups. The literature we draw upon describes qualitative relationships among factors (e.g., that greater similarity leads to greater interpersonal bonds) but rarely defines the precise functional form (e.g., a linear trend or one with diminishing return) or parameter values (e.g., the slope of the linear trend). When theory or empirical evidence was inadequate, we relied on our best judgment to estimate key parameters in the benefit functions or distributions (Sterman, 2002). As we describe next, we included a calibration step in the model development process to adjust these estimates. We also conducted sensitivity analyses to assure that our results are robust when these functional forms or parameter estimates vary within a reasonable range. Finally, we describe the model implementation. We use a movie discussion forum to illustrate how the model works; however, we believe the model applies broadly to any text-based, conversationally oriented online community. In the model, we use the term "member" when describing the motivation and behaviors of people in the real-world, online community and the term "agent" when describing the behaviors of "people" as simulated in the model and represented as decision rules.

4.1. Member Benefits and Costs

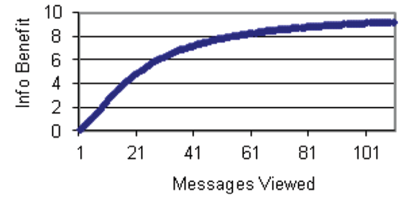
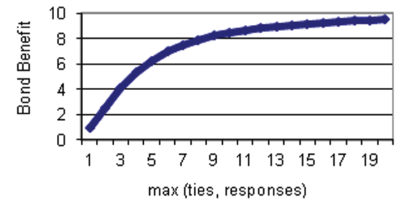
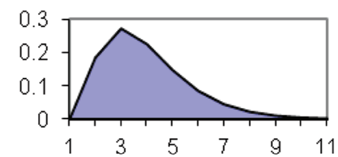
Figure 3 provides an overview of how informational, social, and other benefits are implemented in the model. The table includes the theories used to make assumptions, the rules used to calculate different types of benefits, and key parameters in the benefit functions.

Benefits From Information Exchange

We model two types of benefits related to information exchange: those an agent receives from accessing information and those the agent receives from sharing information with others.

Benefit from Accessing Information. According to the resource-based theory (Butler, 2001), large groups provide more resources—such as content—and therefore greater information benefits to their members. According to information overload theory, however, people have limited capacity to process information. Information overload occurs when there are too many messages or when there is too much noise in the communication (Gu et al., 2007; Jones et al., 2004). In the model, we assume that (a) agents receive information benefit only from messages matching their interests and (b) they receive information benefit as a marginally decreasing function of the number of messages they view. We calculate the benefit from reading messages as a joint function of the quantity and quality of messages an agent reads. On average, the more messages an agent reads that match his interest, the greater information benefit he receives, with diminishing returns due to information redundancy or information overload. The graph in Figure 3 illustrates the information access benefit function. The parameters are based on Liang, Lai, and Ku's (2007) experimental study of

FIGURE 3. Rationale, rules, and functions to calculate member benefits.

	Rationale	Rules	Function and Parameters
Information Benefits			
From accessing information ($InfoB_{access}$)	Information overload (Jones et al., 2004)	Only messages matching an agent's interest provide $InfoB_{access}$, and $InfoB_{access}$ is a marginally decreasing function of the number of messages read.	
From sharing information ($InfoB_{share}$)	Collective effort model (Karau & Williams, 1993)	$InfoB_{share}$ is conditional on liking task or group. It is greater when others undercontribute and when group size is smaller.	
Social Benefits			$f\left(\frac{count\ (viewed\ messages\ that\ match)}{count\ (viewed\ messages)}\right)$
From attachment to group ($SocB_{iden}$)	Group identity (Hogg & Terry, 2000)	$SocB_{iden}$ is greater when an agent's interests are similar to group interests.	
From attachment to members ($SocB_{bond}$)	Interpersonal bonds (Sassenberg, 2002) Empirical studies of *Usenet* groups	$SocB_{bond}$ is greater with repeated mutual interactions, with immediate responses from other members, and is a marginally decreasing function of the max of the two. Similar interests and off-topic messages lead to stronger ties from repeated interactions.	
Intrinsic and Extrinsic Benefits			
From enjoyment ($IntrB_{enj}$)	Intrinsic motivation Empirical studies of online behaviors (Cotte et al., 2006; Ridings & Gefen, 2004)	$IntrB_{enj}$ is a function of individual differences, distributed normally at the agent level, and distributed as a gamma distribution at the community level.	
From reputation ($ExtrB_{rep}$)	Incentive mechanisms (Wasko & Faraj, 2005)	When an individual member contributes at or above 10% of the highest level of contribution	$f\left(\frac{self\ contribution}{max\ contribution/10}\right)$

recommending Internet news articles, which found that an increase from 20 to 40 news items caused information overload and led to reduction in user satisfaction. Compared with news items, messages in online communities are shorter and less complex; thus, we increased the value at which marginal benefit starts decreasing from 20 news items to 40 messages.

Reading messages takes time and effort. We assume that reading cost is proportional to the total number of messages an agent reads. In addition, having to evaluate and discard uninteresting messages increases the cost of reading (Gu et al., 2007). We calculate reading cost as a function that is proportional to the total number of messages the agent views divided by the signal-to-noise ratio (i.e., the number of messages that match the agent's interests divided by the number of messages that fail to match his interests).

Benefit from Sharing Information. In many online communities, a small proportion of members engage in altruistic behaviors, such as answering questions (Fisher, Smith, & Welser, 2006) or performing community maintenance tasks like policing the site (Butler, Sproull, Kiesler, & Kraut, 2007). Engaging in these altruistic actions can lead to positive self-evaluation of competence because it feels good to help others and the community (Wasko & Faraj, 2005). According to the collective effort model (Karau & Williams, 1993), motivation to contribute to group outcomes decreases if members believe that (a) the group is large or (b) others are already contributing, both of which make help less necessary. On the other hand, motivation to share information to help others and the community increases when people perceive group tasks as interesting when they identify with the group, or like other members of the group.

The pseudo code in Figure 4 shows how we implement these rules in the model. If the agent enjoys reading messages or feels strongly attached to the group or its members, we calculate two constituents—one is non-zero when the group is perceived to be at risk of failing (operationalized as hosting fewer than 100 messages), and the

FIGURE 4. Pseudo-code for calculating benefits from sharing information.

```
Initialize information sharing benefit to zero
/* only contribute when task valence or group valence is high */
IF any of intrinsic benefit, identity benefit, bonds benefit >= 3 THEN
   /* more likely to contribute when group is at stake*/
   IF total messages < 100 THEN
      Increase information sharing benefit by 5 times (100 – total messages) / 100
   /*more likely to contribute when perceiving others as under-contributing*/
   IF average other contribution < 10% of self contribution THEN
      Increase information sharing benefit by 3 times self / other contribution
ENDIF
/* less likely to contribute in groups larger than 15*/
IF group size > 15 THEN
   Multiply information sharing benefit by (1 – (group size – 15) / (group size + 15))
ENDIF
```

other is non-zero when others are perceived as undercontributing. We assume that agents who have a history of contributing 10 times more than the community average tend to perceive others as undercontributing and therefore compensate for others' lack of contribution. Finally, to capture the diffusion of responsibility effect, we divide the sum of all constituents by a marginally decreasing function of group size (i.e., the total number of others who are present to contribute).

Benefits From Social Attachment

Prior literature shows that both identification with the group as a whole (i.e., a sense of belonging) and interpersonal bonds with particular members (i.e., friendship) can lead to social attachment and contribution toward group efforts (e.g., Prentice, Miller, & Lightdale, 1994; Sassenberg, 2002). We model identity-based attachment and bond-based attachment separately because they have distinguishable antecedents and consequences (Ren et al., 2007).

Benefit from Identity-Based Attachment. Group identity theory suggests that assigning a member to a group, the presence of an out-group, and similarity among group members all lead to stronger attachment to the group (Hogg & Terry, 2000). Shared interests and similarity in preferences have been used to manipulate and measure identity in laboratory experiments (Amichai-Hamburger, 2005; Postmes & Spears, 2000). We assume that people who share a common interest with the community identify with it. For example, movie lovers feel a stronger sense of belonging to a discussion group if the other members are also movie lovers and if the conversation is about their shared interests instead of other off-topic subjects. In the model, we operationalize benefit from group identity as a function of the similarity between an agent's interest and the community's interest, calculated as the percentage of viewed messages that correspond to the agent's interests: the higher the percentage, the greater the benefit from identity-based attachment.

Benefit from Bond-Based Attachment. Research on small groups indicates that people like each other more as the frequency of their interaction increases (Cartwright & Zander, 1953; Festinger, 1950). Studies of *Usenet* groups suggest that getting a quick reply after posting encourages members of an online community to return and participate in community discussion (Kraut, Wang, Butler, Joyce, & Burke, 2007). Replies from other members signal the likelihood of forming relationships with others in a community. In the model, we calculate the benefit from interpersonal bonds as a function of the number of other agents with whom the agent has developed a relationship through repeated interaction (i.e., the two agents have responded to each other at least twice), weighted by the strength of the relationship and the number of responses the agent received during the last period of interaction, whichever is higher. Research shows attitude similarity (Byrne, 1997) and personal self-disclosure (Collins & Miller, 1994) lead to liking. We therefore assume interacting agents will build stronger relationships if they have similar interests or if the interaction involves off-topic subjects of a particular sort, those revealing aspects of one's self. Benefit from interpersonal bonds has a marginally decreasing form, as illustrated in Figure 3. The first few relationships an agent develops bring greater social benefit than subsequent ones.

Benefit From Enjoyment

Another motivation that leads people to join online communities is the enjoyment they derive from reading and posting online (Ridings & Gefen, 2004) or the intrinsic motivation from engaging in community activities (Deci & Ryan, 2000). Several studies have identified stable individual differences in the extent to which people think online behavior is fun (e.g., Cotte, Chowdhury, Ratneshwar, & Ricci, 2006). For instance, posters enjoy online interaction more than lurkers (Preece, Nonnecke, & Andrews, 2004). Our model captures these individual differences by drawing an agent's interest in reading and posting randomly from a right-skewed gamma distribution (as illustrated in Figure 3). With a gamma distribution, the majority of members have a moderate level of interest in reading and posting messages in online communities and only a small proportion of members have a high level of interest.

Benefit From Reputation

People are also motivated to contribute to online communities by the reputation they gain from doing so (Wasko & Faraj, 2005), signifying a type of extrinsic motivation (Resnick & Zeckhauser, 2002). Many online communities play on this motivation by institutionalizing "leader boards" and other devices that show the most active contributors. Amazon.com, for instance, uses its "top reviewers list" to recognize people who have contributed many reviews. Even when official recognition is absent, active contributors often get recognized by other members as experts in certain topics or as enthusiastic help-providers. In the model, agents who contribute at or above 10% of the highest level of contribution receive reputational benefit. Sensitivity analyses indicate the main results were robust when the proportion of contributors receiving reputational benefit varied between 5% and 15%.

Costs of Participation and Contribution

We model three types of costs associated with reading and posting messages. Access cost simulates the time and effort people spend logging in to read and post messages. Posting cost simulates the time and effort spent composing messages. Compared with reading, posting is more time-consuming and, thus, incurs a higher cost. For simplicity, we assume that starting a new thread and replying to an existing thread incur equal cost. Reading and posting messages also incur opportunity cost, which is the time that could have been spent on alternative activities such as working, conversing with family members, or reading and posting in other communities. We assume that opportunity costs are constant across different online communities but variable across individuals (e.g., opportunity cost is higher for midcareer wage earners than for teens or retirees).

Motivation as a Weighted Sum of Benefits

After the model calculates these benefits and costs, it calculates member motivation as a weighted sum of their benefits from reading and posting minus the costs of reading and posting. All benefits fall within the range of [0, 10] so that we can calculate the weighted sum. These weights differ across communities (Ridings & Gefen, 2004). In the model, we set the weights for information exchange at 0.5, identity at 0.1, bonds at 0.3, and enjoyment and reputational benefits at 0.1, consistent with Riding and Gefen's (2004) findings on interest communities.[3] Within a single community, different members vary in their reasons for participation. Some may go to a movie discussion site to get information, such as movie recommendations, others for companionship with like-minded people, and others because they identify themselves as movie buffs. In the model, weights for individual agents were drawn from normal distributions around the community means.

4.2. Member Actions: Reading and Posting Messages, Entry, and Exit

Calculated motivation determines whether a member reads and posts messages. Following Butler (2001), we define participation as an action members take to be exposed to the community's communication, such as visiting the site to read messages. We define contribution as an action members take to engage actively in community activity, such as posting messages. Following the utility-like logic underlying the expectancy-value theories, we assume a member (a) logs in to read messages when expected benefit from participation exceeds expected cost, and (b) posts messages when expected benefit from contribution exceeds expected cost. Figure 5 provides an overview of the decision rules an agent uses to decide whether to take various actions.

Which Messages to Read?

The model assumes that a member reads messages in reverse chronological order and stops viewing when he runs out of time, interest, or messages. It also assumes that the total number of messages a member views on a particular day depends upon the total number of messages available and the member's motivation to read. The model calculates the number of messages an agent will read on a specific day as proportional to the amount of benefit he received in the past from reading messages minus reading costs, capped by the total number of messages available to read. Because most people read in reverse chronological order and old messages get stale, members are more likely to view and respond to recent messages (i.e., messages posted within a day or so) and have a lower probability of reading older and less active messages (Arguello

[3]Note that in this article we simulate behaviors in an interest community, like a movie discussion group, and do not vary community type. Doing so would involve varying these weights. For example, the relative weights in a technical support group, in which people typically care less about interpersonal bonds and more about information, identity, and reputation, the weights may be 0.5, 0.25, 0.1, and 0.15, respectively. In contrast, the weights for a cancer support group, where one's disease helps defines one's identity, may be 0.33, 0.33, 0.33, and 0.1.

FIGURE 5. Definition and rules for member decisions.

Decisions	Definitions	Rules
Participation	Reading messages	Login and read if expected benefit from reading exceeds expected cost of reading
Contribution	Posting messages	Post if expected benefit from posting exceeds expected cost of posting
Message selection	Which messages to read?	Read newest messages followed by less recent messages, proportional to expected benefit from reading
Topic selection	What is the message topic?	Post topics are jointly determined by topics of recently viewed messages, personal interest, and topic of original message when posting a reply message
Conversation selection	Which message to respond to?	Choosing conversations to join is jointly determined by popularity, reciprocity, and match of personal interest

et al., 2006; Kalman, Ravid, Raban, & Rafaeli, 2006). In deciding whether to post a message, members are influenced by costs, such as whether they must log in or complete a 'CAPTCHA' (Von Ahn, Blum, & Langford, 2004) to contribute. To post a message, an agent makes three additional decisions: (a) whether to start a new thread or to reply to an existing one, (b) which message to respond to if replying, and (c) the topic of a new post if starting a new thread. We assume an agent is equally likely to start a new thread or to reply to an existing thread and sensitivity analyses indicate that our results are robust when the likelihood of starting a new thread varies between 30% and 70%.

What Is the Topic?

A movie discussion community can be organized broadly, welcoming any movie-related topics such as movie genres, critics, and celebrities, or more narrowly around a single topic, such as fantasy movies or Harry Potter films. A member can be interested in one or more of the topics. We assume members' interests remain static and do not change over time. We also assume each message concerns only one topic, although the analysis would be the same if each message encompassed multiple topics. In the model, when an agent starts a new thread, the topic of this message is a joint function of the agent's interests and the topics of the messages the agent has recently viewed. When an agent posts a reply, the topic is a joint function of the topic of the replied-to message, the agent's interests, and topics of the messages the agent has recently viewed. Thus, a fantasy movie lover is likely to initiate or reply to messages about fantasy movies, and this tendency will be greater in a fantasy movie discussion forum than in a general movie forum. In communities with little off-topic discussion, members are less likely to bring up off-topic subjects for fear of violating group norms (Sassenberg, 2002). Theory also suggests that newcomers are more likely to post on-topic messages than old-timers (Ren et al., 2007). Thus, we assume that agents posting for the first time always begin with on-topic messages.

Which Message to Reply to?

Theory and empirical evidence (Faraj & Johnson, 2011; Fisher et al., 2006; Johnson & Faraj, 2005) suggest three common patterns of interaction among community members: (a) preferential attachment, in which members respond to popular messages or posters; (b) reciprocity, in which members respond to others who have written to them in the past; and (c) interest matching, in which members respond to messages that match their interests. Of course, people will respond only to messages they have read. The agent in the model chooses to reply to a message based on a weighted sum of (a) the number of replies the message has received, (b) the number of times the poster of the message has responded to the agent, and (c) the match between the topic of the message and the agent's interests. Agents weigh the three factors equally with a slightly greater weight given to match of interest in technical communities, to reciprocity in interest and support communities, to reciprocity and popularity in political discussion communities (Turner, Smith, Fisher, & Welser, 2005). Sensitivity analysis shows that our results are the same whether we assume equal or unequal weights.

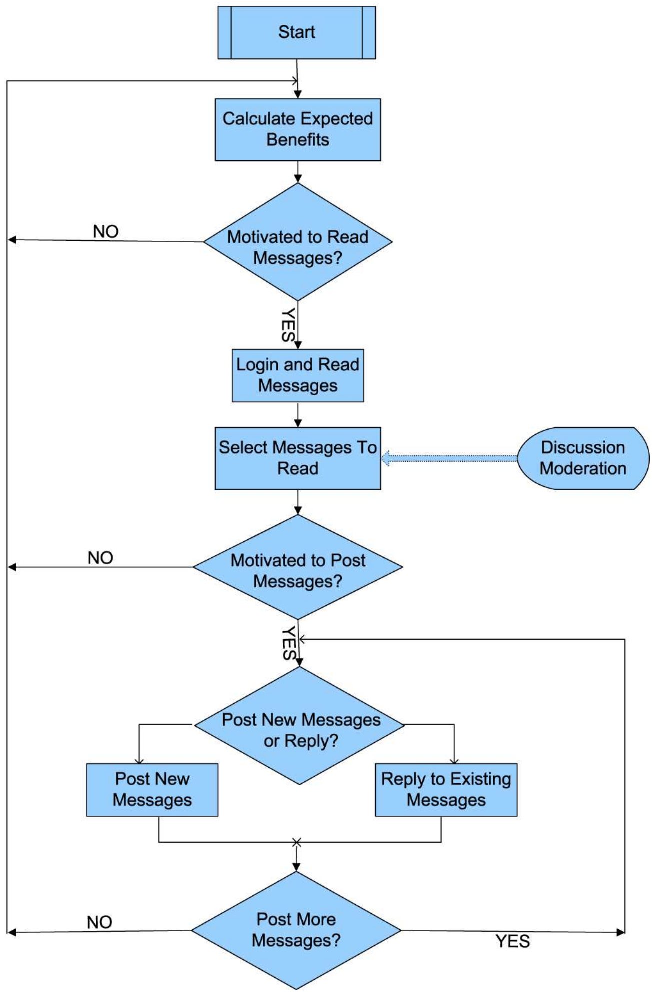
Member Entry and Exit

Members join and leave online communities. Because there is little prior research describing the rate at which newcomers enter online communities, we analyzed 100 Usenet groups to estimate the parameters related to entry. This analysis indicates that the number of newcomers joining a community is proportional to community size (see Butler, 2001, for similar results) and follows a truncated gamma distribution function. Larger communities attract more newcomers per day. Typically, a community attracts a mean number of newcomers on most days and attracts an atypically large number of newcomers on a small number of days. In the model, agents become inactive if they have not posted for 30 days and have not visited the community for a substantial period, like a year. The change of status is stochastic meaning the more days they have not visited the community, the greater likelihood of their status being changed to inactive.

4.3. Model Implementation and Calibration

We implemented the model using *NetLogo*, a cross-platform multiagent modeling environment (Wilensky, 1999). Within the simulation, agents take actions during a simulated period. In our simulation the period is a day, because this is the temporal granularity we had available from empirical data, although the logic of the simulation is similar if we shorten the period to 1 hr or 1 min. Figure 6 depicts the sequences of agent decisions. All active agents decide to read and post before anyone moves to

FIGURE 6. Sequences of decisions an agent makes in a simulated day.



the next period. Messages posted in the previous period are made available to be read by all agents in the next period, and whatever messages an agent reads are used to update their expectations of benefits. In the jargon of agent-based modeling, actions are organized in staged episodes, and time is simulated as forced parallel.

We took three steps to ensure the external validity of the model. Whenever possible, we drew insights from existing theories to specify the key assumptions and relationships in the model. The prior sections describe this rationale. When theory was insufficient, we mined data from 100 Usenet groups to fix important parameters such as the ratio of new threads to replies or the entry rate for newcomers. We also went through an iterative calibration process during which we systematically varied key parameters to replicate behavioral patterns that have been repeatedly discovered in empirical studies, such as the power-law distributions of posts per members. We describe this calibration in more detail next. To assure the robustness of our results, we ran sensitivity analyses by varying a selective set of parameters. Results do not differ substantially from those we report next. The sensitivity analyses were by no means exhaustive; instead, we focused on parameters that were likely to alter the results.

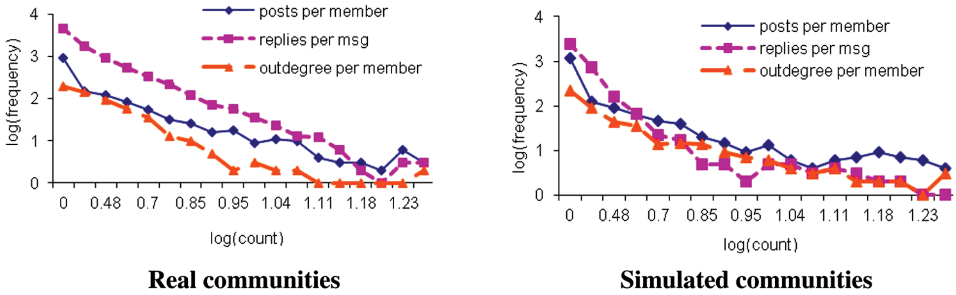
Model calibration is the process of adjusting a computational model to produce results that match real data or stylized facts with reasonable accuracy (Carley, 1996). We calibrated our model so that it reproduced the power-law distribution of three statistics, which prior research has shown characterize online communities—posts per member, replies per post, and communication partners (out-degrees) per member (Fisher et al., 2006; Smith, 1999). The process involved tweaking parameters so the model generated simulated data that matched training data from 12 Usenet groups. We then validated this calibrated model against data from a new sample of 25 Usenet groups. We used pattern calibration to establish the reasonableness of the model and its potential for predictive accuracy. Pattern calibration compares the pattern or distribution of results generated by a computational model with the pattern or distribution generated from real data.

We first simulated 12 online groups starting with data on membership size and message volume from 12 real Usenet groups.[4] The groups started with 30 to 500 posters and 30 to 500 active messages at the beginning of the simulated period. During the calibration phase, we engaged in an iterative process in which we compared the distribution of the three statistics from the simulation—posts per agent, replies per post, and out-degrees per agent—with data from the real groups. After each run, we examined mismatches between the simulated and the real data, reexamined assumptions, and made adjustments to the model in light of theoretical reasoning, empirical evidence, or knowledge about how the processes in the model operate. After 10 iterations, the model replicated the power-law distribution for all three statistics. The iterative calibration process helped select parameters, variables, and relations that yield outcomes that correspond to the real world (Burton & Obel, 1995), which greatly increases the construct and external validity of our model.

During the validation phase, we simulated another 25 online groups, starting with data on membership size and message volume from real groups. The simulated statistics fit the actual statistics for these 25 real groups well and demonstrate the validity of the model. Figure 7 illustrates the distribution of the real and simulated statistics (after log transformation) in one of these 25 groups. We calculated the Pearson correlations between the real data series and the simulated data series, as shown in Figure 7, for all three statistics. The average correlation across the 25 groups between the empirical and simulated data ranged between 0.90 and 0.96, confirming that the model matched behavior in the hold-out sample well. We also examined survival curves for members and messages during model calibration and validation, and also found high similarity between simulation and real data. As shown in Figure 8,

[4]The Usenet groups in the data set had no moderation, which is the condition we used for model validation.

FIGURE 7. Comparison of descriptive statistics from real and simulated communities.



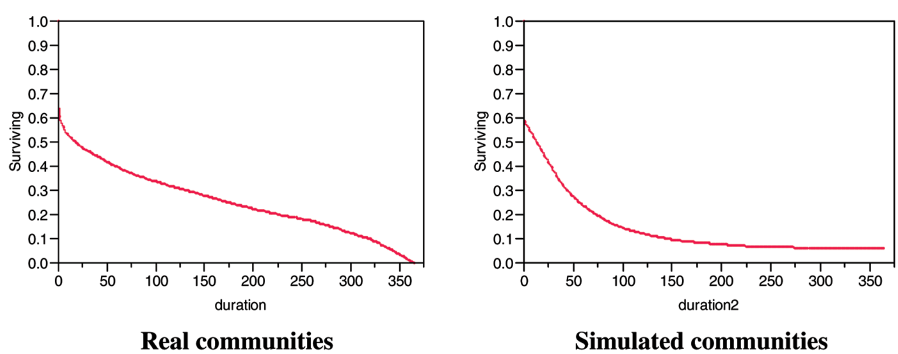
the survival curve from real data indicates that about 60% of new posters fail to return after their first post and on average only 10% to 20% remain active for more than 100 days.

5. SIMULATION EXPERIMENTS AND RESULTS

5.1. Virtual Experimental Design

In this section, we describe a full-factorial simulation experiment varying three parameters over which a community designer has some control: the number of topics a community officially supports, its message volume, and the type of moderation it uses. The experiment simulated three levels of topic breadth by populating the community with agents having interests in one, five, or nine topics. It simulated three levels of message volume, with an average 10, 15, or 20 messages per day, by varying the costs of sending messages. It simulated three types of moderation: no moderation, community-level moderation (which removed messages that do not conform to nominal topics—i.e., off-topic messages), and personalized moderation (which showed agents subsets of messages that matched their interests).

FIGURE 8. Comparison of member survival from real and simulated communities.



Imagine a movie discussion forum with five germane topics (upcoming movies, upcoming DVDs, movie critics, celebrities, and video games) and an agent interested in two of these (upcoming movies and celebrities). In a community with no moderation, the agent sees all messages—regardless of topic—with the most recent first. In a community with community-level moderation, all agents see only on-topic messages, because a moderator removes messages that do not correspond to topics supported in the community. Under personalized moderation, the agent sees only messages that match his interests (in this case, messages about upcoming movies and celebrities). In addition, the agent can also choose to view off-topic messages, for example, about places for vacation, if he finds them interesting. Filtering occurs at the individual level and other agents will see a different selection of messages.

We ran two 365-day simulations for each condition, one with five randomly constructed groups and one with 20 groups. The results were similar in the two replications. We report here results from the five-group simulation. All groups began with 30 seed agents and 30 seed messages and evolved over time as newcomers joined and old-timers left. On each simulated day, each agent assessed prior benefits from having read and posted messages and decided whether to log in to read and post new messages. For purposes of the simulation, the accuracy of the personalized moderation was set to 80% of recommended messages matching a member's interests. Sensitivity analyses suggest our results remain robust when the accuracy of personalized moderation varies between 60% and 100%.

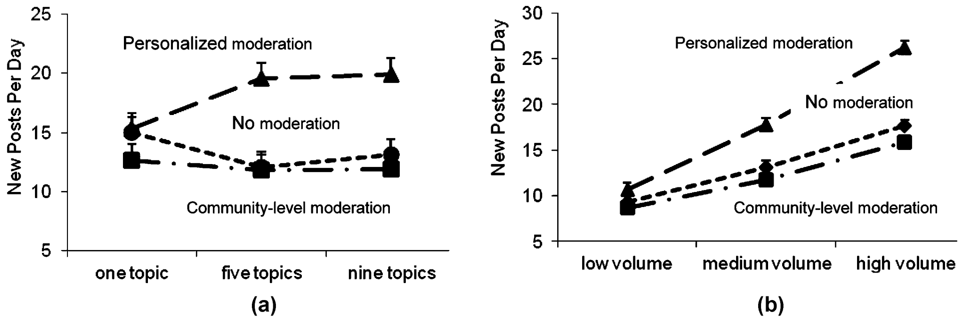
We examined the effects of topical breadth, message volume, and moderation on two outcomes easily visible to any community manager: the number of new posts per day, which is an indicator of community activity, and the average number of login sessions per member, which is an indicator of member commitment. We ran an analysis of variance (ANOVA) to examine the effects of moderation, topical breadth, and message volume on the two outcomes. We also examined the benefits members received at the 100th, 150th, 200th, 250th, and 300th day of the experiment—a total of five snapshots. We analyzed informational and relational benefits, which our simulation results revealed as mediators of community activity and member commitment, to understand the link between the three design decisions and community outcomes.

5.2. Simulation Results

Effects on Community Activity

Analyses of *community activity* in communities differing in *topical breadth* and *moderation* revealed no significant effects of topical breadth ($p = .83$), a significant main effect of moderation ($p < .001$), and a significant interaction between the two ($p = .05$). As shown in Figure 9a, personalized moderation led to the highest level of community activity (15 to 20 posts per day), which was about 50% more than community-level moderation and 36% more than no moderation, with no significant difference between the latter two ($p = .24$). Figure 9 also shows that personalized moderation was especially beneficial in communities with greater topical breadth. In communities

FIGURE 9. Effects of design decisions on community activity. (a) Topical Breadth. (b) Message Volume.



supporting five or nine topics, personalized moderation led to approximately 65% more posts than community-level moderation and no moderation.

Analyses of *community activity* in communities differing in *message volume* and *moderation* revealed two main effects and a significant interaction between moderation and message volume ($p < .001$). By definition, communities with high message volume had more posts per day than communities with low volume (20 messages vs. 10 messages). Personalized moderation, again, led to more posts than community-level moderation and no moderation, and the difference was much greater in communities with higher message volume ($p < .001$). As shown in Figure 9b, personalized moderation led to 49% to 66% more posts than no moderation and community-level moderation in communities with high message volume versus to 14 to 22% more posts in communities with low message volume.

Effects on Member Commitment

Analyses of *member commitment* in communities differing in *topical breadth* and *moderation* revealed a significant main effect of topical breadth ($p = .01$), a main effect of moderation ($p < .001$), and a significant interaction between the two ($p < .001$). As shown in Figure 10a, members of communities with moderate to

FIGURE 10. Effects of design decisions on member commitment. (a) Topical Breadth. (b) Message Volume.

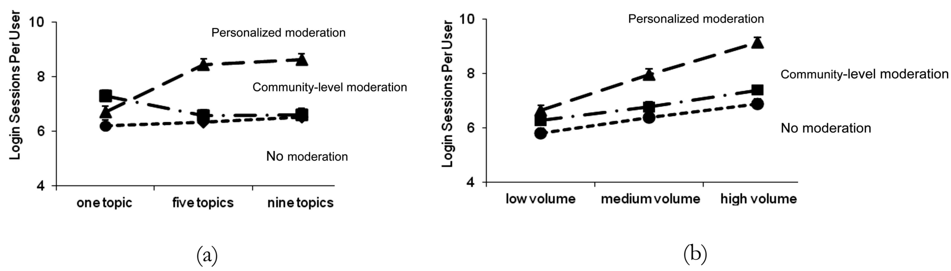
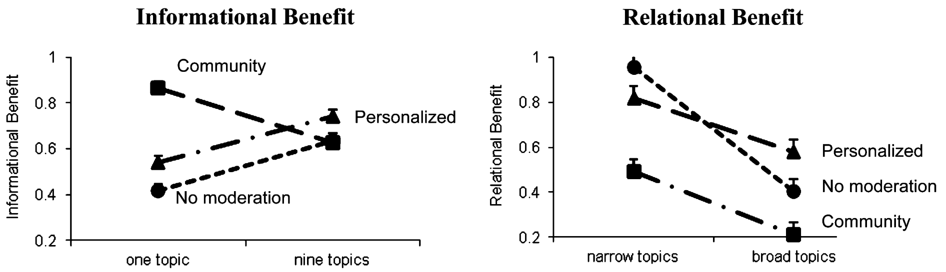


FIGURE 11. Effects of moderation on informational and relational benefits by topical breadth.



high topical breadth visited more frequently than members of communities with a narrow focus. Both personalized and community-level moderation led to more frequent visits than no moderation but under different conditions. Community-level moderation led to the highest login frequency in communities with a single topic, whereas personalized moderation led to the highest login frequency in communities with more topics.

The interaction of *message volume* and *moderation* on commitment is similar. Figure 10b shows that higher message volume led to more frequent logins—from approximately six logins in communities with low volume to eight logins in communities with high volume. Personalized moderation led to more frequent logins than community-level moderation ($p < .05$), and community-level moderation led to more frequent logins than no moderation ($p < .05$). Compared to no moderation, personalized moderation to more frequent logins in communities with medium and higher message volumes than in communities with low message volume ($p = .01$, 33% vs. 14% increase).

5.3. Member Benefits From Information and Interpersonal Bonds

Posts and logins are observable behaviors. In the simulation, posts and logins are driven by the benefits that agents have received in the past. To better understand the mechanisms through which design decisions affect posts and logins, we examined the impact of the design decisions on two benefits that preliminary analysis suggested were especially important: benefit from accessing information (informational benefit) and benefit from interpersonal bonds (relational benefit).

5.4. Member Benefits in Communities with Different Topical Breadth[5]

Figure 11 shows the effects of moderation and topical breadth on informational and relational benefits. On average, agents received more *informational benefit*

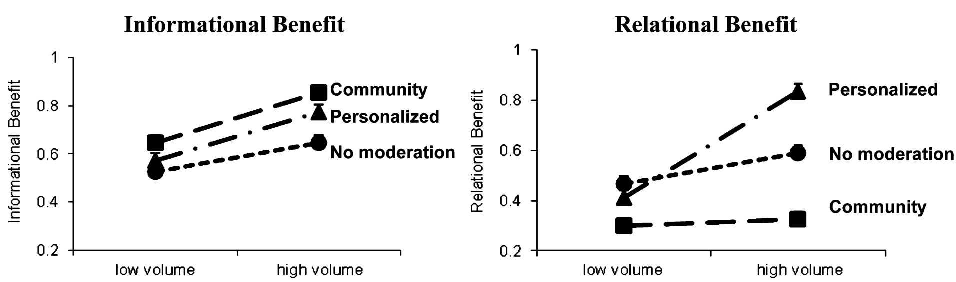
[5]Because previous analyses revealed no significant difference between medium and broad topical breadth and a linear effect of message volume, we omitted medium topical breadth in Figure 11 and medium message volume in Figure 12 to make the figures more readable.

in topically broad communities than topically narrow ones ($p < .01$), and with either type of moderation than with no moderation ($p < .001$). Community-level moderation led to twice as much informational benefit compared to personalized or no moderation in communities with a narrow focus, whereas personalized moderation led to 10% to 15% greater benefit in communities with a broad focus ($p < .001$). In contrast, agents received greater *relational benefit* in the topically narrow communities than in the topically broad communities ($p < .001$). Both personalized moderation and no moderation led to greater relational benefit than did community-level moderation ($p < .001$). The effects of moderation on relational benefit also depended on topical breadth ($p < .001$). Compared to no moderation, personalized moderation led to a 15% decrease in relational benefit in communities with a narrow focus but a 43% increase in relational benefit in communities with a broad focus.

5.5. Member Benefits in Communities With Different Message Volume

Figure 12 shows the effects of moderation and message volume on informational and relational benefits. Agents received greater *informational benefit* in communities with higher message volume ($p < .001$) and in communities with either personalized or community-level moderation than no moderation ($p < .001$). Compared to no moderation, personalized moderation led to a 23% increase in informational benefit, whereas community-level moderation led to a 34% increase. The interaction between moderation and message volume was not significant ($p = .10$), suggesting that the effect of moderation on information benefit did not vary substantially in communities with different levels of message volume. Similarly, agents experienced greater *relational benefit* in communities with higher message volume ($p < .001$). Personalized moderation led to the greatest relational benefit, followed by no moderation and community-level moderation ($p < .001$). There is significant interaction between moderation and message volume ($p < .01$). As message volume increases, the effects of no moderation and community-level moderation remained roughly the same, whereas the effects of personalized moderation doubled.

FIGURE 12. Effects of moderation on informational and relational benefit by message volume.



6. DISCUSSION

6.1. Summary of Findings

Figure 13 summarizes our main findings. The dynamic, nonlinear, and interactive nature of agent-based modeling implies that we can only speculate, after the fact, what might have caused these results (Carley, 2002). The most robust finding from the simulations is the superiority of personalized moderation in increasing both members' commitment and contribution, with the effects larger in more topically diverse and higher volume communities. This pattern can be explained by trade-offs between informational and relational benefits. Providing a one-size-fits-all view for all members—with either no moderation or community-level moderation—makes it difficult to meet the needs of members with different preferences for various benefits. Personalized moderation resolves the trade-off by customizing a selective set of messages to match members' preferences, whether they be informational, relational, or both. Supporting a broad range of topics and high message volume increases the challenge of balancing diverse preferences, which is why personalized moderation has stronger effects in communities with these conditions. On the other hand, personalized moderation comes with some costs as well. Personalized moderation significantly reduces the number of messages shown to a member. With an accuracy rate of 60% or 80%, personalized moderation might miss relevant messages while including some irrelevant ones. However, its positive effects suggest most users prefer to see less volume but content of higher quality.

A second interesting finding is that members of topically broad communities are more committed or tend to visit more frequently than members of topically narrow communities, although they do not post more messages. This result can be partially explained by the effects of topical breadth on informational and relational benefits. On one hand, having more topics to discuss increases informational benefits because it increases the number of messages likely to match one's interest. On the other hand, it reduces relational benefit because it reduces the chance that two randomly matched members will share a common interest. One of the model's assumptions based on theories of interpersonal bonds is that in repeated interactions, members with similar interests develop stronger relationships with others in their community than do members with dissimilar interests. The effects of topical breadth depend upon the interplay between all benefits, including informational and relational ones. Greater topical breadth increases visit frequency but not number of postings, probably because posting incurs greater costs than reading, and the net benefit from broad topics is sufficient to overcome the cost of reading but not the cost of posting messages.

The third finding is that community-level moderation leads to greater commitment but not contribution. One possible reason is, again, the differential effects of moderation on informational and relational benefits. As mentioned earlier, community-level moderation removes off-topic messages. Although off-topic messages, such as personal stories, do not provide informational value, they can

FIGURE 13. Summary of findings from simulation experiments.

Design Decisions	Community Activity (# Posts)	Member Commitment (# Logins)	Informational Benefit	Relational Benefit
Topical breadth	No significant effects	More frequent logins in topically *broad* communities	*Greater* informational benefit in topically *broad* communities	*Reduced* relational benefit in topically *broad* communities
Message volume	More posts with higher message volume	More frequent logins with higher message volume	Greater informational benefit with high volume	Greater relational benefit with high volume
Community-level moderation	No significant effects	More frequent logins than no moderation in topically *narrow* communities	*Greater* informational benefit in topically *narrow* communities	*Reduced* relational benefit than other moderation in all communities
Personalized moderation	More posts than other types of moderation, especially in topically *broad* communities and *high* message volume	More frequent logins than no moderation in topically *broad* communities and with *high* message volume	Greater informational benefit, especially in topically *broad* communities	Greater relational benefit especially in topically *broad* communities and with *high* message volume

augment the process of relationship development by helping people get to know each other (Collins & Miller, 1994). One of the assumptions in our model is that off-topic messages provide opportunities for personal disclosure and lead to stronger interpersonal bonds than do messages of nominal topics. By removing off-topic messages, community-level moderation increases informational benefit at the price of reducing relational benefit. Again, because posting messages incurs greater costs than reading messages, the net benefit from community-level moderation might be sufficient to encourage more frequent logins to read messages but not more frequent posting. Another possible explanation is that community-level moderation may have affected the retention of posters and lurkers differently. Due to the public goods nature of online conversations, posters and lurkers have equal access to information provided by other members. By limiting off-topic messages, community-level moderation may provide disproportionate benefit to lurkers, who are driven primarily by informational benefit, causing them to return more frequently, rather than to posters, who are driven by both informational and relational benefits.

6.2. Contributions to Online Community Literature

In this article, we synthesized multiple social science theories in an agent-based model to develop new theories to understand trade-offs in online community design. By integrating propositions from multiple theories, our model depicts a more complete picture of how individual motivation and interactions affect community dynamics than any of the constituent theories can. Our effort makes four contributions to the human–computer interaction literature in general and online community literature in particular.

First, we contribute a systematic understanding of the pros and cons of three design decisions regarding topical breadth, message volume, and styles of discussion moderation in the context of building a vibrant online community. Our application of the agent-based model to understand how three design decisions affect community activity and member commitment led to plausible yet nonobvious predictions, such as the differential effects of topical breadth and community-level moderation on member commitment versus contribution. These results call for reconsideration of well-established beliefs in the effectiveness of a narrow focus and community-level moderation. Both were shown to be less effective than either their common use or experts' opinions would imply. For instance, Preece (2000) notes a moderator's number one task is to "keep the group focused and on-topic" (p. 84). In contrast, our study suggests that a narrow focus promotes relational benefits at the expense of informational benefits, whereas community-level moderation is effective only in narrowly defined communities and improves informational benefits at the expense of relational benefits.

Second, the measurement of benefits that intervene between design decisions and observable outcomes enabled us to examine not only the end results of the design decisions but also how they were produced, illustrating a trade-off between informational and relational benefit. We call it a trade-off because design decisions, such as a narrow topical focus or community-level moderation, tend to promote one type of benefit (e.g., informational) at the cost of the other (e.g., relational). The trade-off occurs partially because informational and relational benefits originate from different factors. For instance, informational benefit accrues from on-topic messages that match one's interest whereas relational benefit accrues from off-topic messages that disclose personal information about the poster. Several of our findings can be attributed to the interplay between informational and relational benefits. This illustrates an advantage of using agent-based modeling to integrate simple social theories. Although most theories focus on a small set of factors influencing a single outcome, our agent-based model predicts the overall effects of design decisions on multiple outcomes simultaneously.

Third, we also contribute a contingency view of online community design. There is no universally optimal design for all communities. For example, each of the three moderation styles can be a good choice, depending upon community characteristics and specific goals designers wish to accomplish (e.g., to make members loyal or to increase their contribution). When the risk of information overload is low, as in small, narrowly defined communities, the community thrives with no moderation or community-level moderation. As communities grow, attracting more members and accommodating a broader set of interests, community designers should consider personalized filtering to reduce information overload and assist members in finding similar others to engage in interesting conversations.

Our final contribution is the agent-based model itself, which represents an extensible theory to understand both why online communities operate as they do and to inform other aspects of online community design. Because traditional social science theories focus on a small set of variables, they do not necessarily fit together into a comprehensive view, nor can they readily be used for prescriptive purposes to make design recommendations (Baligh, Burton, & Obel, 1996). To create an accurate and useful platform that informs online community theory and design, one must put together pieces from constituent theories into a consistent and comprehensive whole. Our agent-based model serves such a purpose.

The model can be extended by incorporating more complete theories of how benefits are produced by incorporating other benefits or by applying it to other decisions in online community design. The model could be extended to make the rules describing each benefit more complete. For example, the rules for producing identity-based social benefits concentrate on the alignment between an agent's interests and the group's norm. It would be possible to incorporate other factors that increase commitment, such as comparing an in-group to out-groups or having members work on a common task (Gaertner et al., 2000). Second, the model is now limited to the six benefits. Other benefits, such as satisfying a need for power or skill development, could be incorporated; for example, many people participate in open source development projects to learn new skills (Lakhani & von Hippel, 2003). The model can also be extended to examine other decisions, such as introducing a leader board and institutionalizing ways for newcomer socialization.

6.3. Implications for Online Community Design

It is common for online community designers and managers to ask questions such as "How broadly should we define the niche of the community?" "How much traffic can we expect?" or "Should there be moderation and what kind of moderation is most appropriate for the community?" Our findings provide some guidelines for making these decisions. Designers may consider a narrowly focused online community with community-level moderation if they believe (a) the narrow focus can attract a critical mass of users without an overwhelming number of messages (e.g., more than 20–30 per day), (b) user preference for informational benefits outweighs their preference for relational benefits, and (c) it is too costly to implement personalized moderation. For example, these factors may characterize a Linux technical support group where members primarily want information on topics like how to install the operating system on new hardware or how to configure the graphic user interface.

Choosing a moderation style becomes more subtle and difficult when expected users have diverse preferences (i.e., to exchange information and form relationships or get social support), as in a cancer support group, where members want personal connections and emotional support in addition to health information. If designers believe (a) they can attract enough members with a narrow focus, (b) members expect both informational and relational benefits, and (c) it is too costly to implement an advanced system of personalized moderation, then they should consider simple ways to implement personalized moderation, such as providing built-in filters or structures (e.g., e-mail filters or subforums) that enable members to see only messages likely to interest them. If online community designers expect to attract users with diverse interests and a relatively high volume of messages, they should include the expense of personalized moderation in their planning effort and explore advanced systems and algorithms for implementation.

Even though personalized recommendation has been widely used to provide people with customized views of digital content, as in Google News and Amazon.com, it has been rarely used to recommend online conversations. Community managers can use personalized recommendations to create clusters of members with similar interests, allowing them to choose to experience a narrow or wide range of information according to their preferences, and then return to their intimate circles for personal conversations. Modern platforms such as Twitter and Facebook filter conversations based on both topics and social ties (Chen, Nairn, Nelson, Bernstein, & Chi, 2010) and users have found these novel algorithms to be more efficient and enjoyable ways of joining and navigating online conversations (Bernstein et al., 2010).

Finally, our agent-based model can be expanded and augmented with a rich user interface to serve as a decision-making tool for community designers. By changing parameters, designers can run "what-if" experiments to navigate the design space and explore different scenarios, matching design decisions to the context of the community (Baligh et al., 1996). The dynamic nature of the model allows community designers to foresee not only the immediate consequence of their decisions but, more importantly, long-term consequences for the health and development of the community in terms of membership base, member compositions, and level of activities. The potential of simulation models to aid real-life decision making is even less recognized than its potential for theory development. Hence, the widespread use of these tools requires persistent effort to refine and customize them for practical use, increasing awareness about their potential and documenting their successful implementation in case studies.

6.4. Limitations

This research is not without limitations. Any agent-based model is a compromise between simplicity and accuracy. To make our model clear and interpretable, we made simplifying assumptions to capture the essence of people's motivations to participate in groups. In this section, we acknowledge these limitations, speculate how altering our assumptions may change our results, and discuss ways to relax these assumptions and extend the model in future research.

We simulated one type of community: a conversation-based one organized around a set of shared interests or topics, such as a movie discussion group. Many other types of communities exist, such as technical and health support groups, political discussion groups, online gaming communities, open source software development projects, and social networking sites. In part, these communities differ in the extent to which members are motivated by different kinds of benefits and the ways in which people interact. For example, members of technical support groups may be less motivated by social benefits and more by informational and reputational benefits than members of interest groups; and members of gaming communities interact by engaging in joint activities as well as by talking. We believe that our main findings—that is, the superiority of personalized moderation and the trade-off between informational and relational benefits—apply to a broad range of conversation-based communities. Yet generalization needs to be done with caution, and future research should examine these effects in all types of communities.

We assumed that members' preferences and interests reflect stable, individual differences. In real communities, however, newcomers who join to talk about the nominal topics may, after repeated encounters, increasingly value friendship with other community members. Likewise, members' interests or attitudes toward certain topics may shift over time in response to the messages to which they are exposed. In developing the simulation, whenever possible, we drew insights from social science theories and empirical evidence of Usenet groups to ground the model. Because of the lack of research on low-level functions and parameters, we had to rely upon our judgment to estimate key parameters in the benefit functions or distributions (Sterman, 2002). We ran sensitivity analyses to ensure our main findings are robust and not dependent upon the values of key parameters. In our model description, we mentioned some of the parameters we varied in the sensitivity analyses, and a complete list is available upon request. This is a limitation of our study and an opportunity for future research. For instance, we assume that bond-based benefit can originate from either repeated interactions or immediate responses from other members and the two are somewhat interchangeable. Are they? This assumption, and many others, can be tested empirically or virtually by systematically varying the functions and parameters in the model.

Although the model presented here was based on a selection of social science theories relevant to motivation in online communities, it was not exhaustive. We identified six benefits that motivate people to participate in online communities and the theories we chose are representative of those that produce those benefits. However, they do not encompass all the motivations that cause people to participate in online communities. For example, vandals, trolls, and others who attempt to disrupt online communities are motivated by needs for control that is not simulated in the model. Nor does the model encompass all the theories relevant to a particular benefit. For example, psychologists have studied interpersonal attraction for decades (see Fehr, 2008, for a recent review). Our modeling of interpersonal bonds does not include all factors known to drive friendship formation such as status, the ability of one person to reward another, social network effects like triadic closure, and physical attractiveness.

Because we were building an agent-based model, the theories we drew upon focused on the motivations and behavior of an individual member. However, we did not draw upon theories that involved other units of analysis, such as cliques within communities or the community as a whole. This is not necessarily a limitation of our study because building the agent-based model requires careful definition of its boundary conditions, both in theory and in practice. Although we believe we chose a useful set of theories, social science literature offers a wide range of other theories that could be exploited to extend the model to examine other types of online communities or other decisions in online community design. One extension could be to include an ecological view, which simulates the process through which users compare potential benefits in the target community versus from other communities (X. Wang, Butler, & Ren, 2013, as a reference). The opportunity cost in the current model acts as a placeholder that can be expanded to incorporate this process. Another extension could be to include new activities to simulate other types of online communities, for example, the creation of *Wikipedia* articles to simulate production-oriented communities. Social science provides a rich theoretical basis such as goal-setting theories, task interdependence and coordination theories, social influence, and collective action theories, just to name a few.

Yet another extension could include financial considerations, both in terms of the revenue from and the expense of hosting and moderating an online community. Many online communities are launched with the expectation to monetize the effort in the future. We drew the boundary to exclude financial considerations from the current model, although the model could be extended to include expected revenue and cost as factors designers consider when making design decisions. The extension requires assumptions about the business model of the community, which is beyond the scope of this article.

7. CONCLUDING REMARKS

Online communities are successful to the extent that members return repeatedly and contribute materials that others value and to the extent that members receive benefits when they visit. Because many decisions are not motivated by a systematic understanding of member motivation and contribution but driven by intuition and trial and error, communities are often less successful than they could be. In this study, we treat online communities as socio-technical systems that need to be carefully designed to fit their strategic goals. In other words, we believe online community design can go beyond intuition and trial and error and can benefit from the prescriptive power of social science theory. We believe agent-based models, which can be used to incorporate multiple theoretical perspectives, have the potential to evolve into a multicontingency tool for diagnosis and design of online communities (Burton & Obel, 2004), and we hope that the agent-based model described here illustrates this potential. Theoretical knowledge and predictions embedded in the model can be combined with creative design intuition to generate effective design decisions. We acknowledge that intuition and trial and error will continue to be essential to the design of online communities; however, we hope that our model can serve as a test bed to help designers gain preliminary knowledge of the features with which they can experiment. We also hope our research on the application of theory to the problem of online community design serves as a case study of how to extract value from social science theories to inform design.

NOTES

Acknowledgments. Earlier versions of this article have been presented at the research colloquiums of Carnegie Mellon University, McGill University, University of Maryland, and University of Minnesota and workshops at Academy of Management Annual Meeting and ACM SIGCHI Conference on Human Factors in Computing Systems. We thank our participants for their valuable feedback. We also thank Shawn Curley for his helpful comments and Sam Hashemi for research assistance.

Funding. This work was supported by Carlson School of Management Dean's Small Research Grant, National Science Foundation Grant IIS-0325049 (Designing Online Communities to Enhance Participation), IIS-0729286 (Solving Critical Problems in Online Groups), and IIS-0808692 (Understanding Online Volunteer Communities). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies of the National Science Foundation.

Supplemental Data. Supplemental data are available for this article. RenKraut-2013-SimulateOnlineCommunity.nlogo is the simulation model and RenKraut-2013-SupplementaryMaterials.pdf provides instructions for downloading the simulation platform, running the model, and interpreting the results. Free access to these materials is available at the online edition of this article on the *Human–Computer Interaction* publisher's website (www.tandfonline.com/hhci).

HCI Editorial Record. First manuscript received August 22, 2012. Revision received June 26, 2013. Accepted by Peter Pirolli. Final manuscript received June 26, 2013. — *Editor*

REFERENCES

Allen, D. (1988). New telecommunications services: Network externalities and critical mass. *Telecommunications Policy, 12*, 257–271.

Amichai-Hamburger, Y. (2005). Internet minimal group paradigm. *CyberPsychology & Behavior, 8*, 140–142.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist, 51*, 355–365.

Arguello, M., Butler, B. S., Joyce, E., Kraut, R. E., Ling, K. S., & Wang, X. (2006). Talk to me: Foundations for successful individual-group interactions in online communities. *Proceedings of the CHI 2006 Conference on Human Factors in Computer Systems*. New York: ACM.

Baligh, H. H., Burton, R. M., & Obel, B. (1996). Organizational consultant: Creating a useable theory for organizational design. *Management Science, 42*, 1648–1662.

Bernstein, M. S., Suh, B., Hong, L., Chen, J., Kairam, S., & Chi, E. H. (2010). Eddi: Interactive topic-based browsing of social status streams. *Proceedings of the UIST 2010 Conference on Human Factors in Computer Systems*. New York: ACM.

Burton, R. M., & Obel, B. (1995). The validity of computational models in organization science: From model realism to purpose of the model. *Computational and Mathematical Organization Theory, 1*, 57–71.

Burton, R. M., & Obel, B. (2004). *Strategic organizational diagnosis and design: The dynamics of fit, with OrgCon Software* (3rd ed.). Boston, MA: Kluwer Academic.

Butler, B. S. (1999). *The dynamics of cyberspace: Examining and modeling online social structure* (Unpublished doctoral dissertation). Carnegie Mellon University, Pittsburgh, PA.

Butler, B. S. (2001). Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research, 12*, 346–362.

Butler, B. S., Sproull, L., Kiesler, S., & Kraut, R. E. (2007). Community effort in online groups: Who does the work and why? In S. P. Weisband (Ed.), *Leadership at a distance: Research in technologically-supported work* (pp. 171–194). Mahwah, NJ: Erlbaum.

Byrne, D. (1961). Interpersonal attraction and attitude similarity. *The Journal of Abnormal and Social Psychology, 62*, 713.

Byrne, D. (1997). An overview (and underview) of research and theory within the attraction paradigm. *Journal of Social and Personal Relationships, 14*, 417–431.

Carley, K. M. (1996). *Validating computational models* (Working paper). Pittsburgh, PA.

Carley, K. M. (2002). Computational approaches to sociological theorizing. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 69–83). New York, NY: Kluwer Academic.

Cartwright, D., & Zander, A. (1953). Group cohesiveness: Introduction. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory* (p. 81). Evanston, IL: Row Peterson.

Chen, J., Nairn, R., Nelson, L., Bernstein, M., & Chi, E. H. (2010). Short and tweet: Experiments on recommending content from information streams. *Proceedings of the CHI 2006 Conference on Human Factors in Computer Systems*. New York: ACM.

Collins, N., & Miller, L. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin, 116*, 457–475.

Cotte, J., Chowdhury, T. G., Ratneshwar, S., & Ricci, L. M. (2006). Pleasure or utility? Time planning style and web usage behaviors. *Journal of Interactive Marketing, 20*, 45–57.

Davis, J. P., Eisenhardt, K. M., & Bingham, C. B. (2007). Developing theory through simulation methods. *Academy of Management Review, 32*, 480–499.

Deci, R. M., & Ryan, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54–67.

Faraj, S., & Johnson, S. (2011). Network exchange patterns in online communities. *Organization Science, 22*, 1464–1480.

Fehr, B. (2008). Friendship formation. In S. Sprecher, A. Wenzel, & J. Harvey (Eds.), *Handbook of relationship initiation* (pp. 29–54). New York, NY: Taylor & Francis.

Festinger, L. (1950). Informal social communication. *Psychological Review, 57*, 271–282.

Figallo, C. (1998). *Hosting Web communities: Building relationships, increasing customer loyalty, and maintaining a competitive edge*. New York, NY: Wiley.

Fisher, D., Smith, M., & Welser, H. T. (2006). You are who you talk to: Detecting roles in Usenet newsgroups. *Proceedings of the 39th Hawaii International Conference on System Sciences*. IEEE Computer Society.

Gaertner, S. L., Dovidio, J. F., Banker, B. S., Houlette, M., Johnson, K. M., & McGlynn, E. A. (2000). Reducing intergroup conflict: From superordinate goals to decategorization, recategorization, and mutual differentiation. *Group Dynamics, 4*, 98–114.

Gu, B., Konana, P., Rajagopalan, B., & Chen, H. W. M. (2007). Competition among virtual communities and user valuation: The case of investing-related communities. *Information Systems Research, 18*, 68–85.

Harper, F. M., Frankowski, D., Drenner, S., Ren, Y., Kiesler, S., Terveen, L., . . . Riedl, J. T. (2007). Talk amongst yourselves: Inviting users to participate in online conversations. *Proceedings of 12th International Conference on Intelligent User Interfaces*. New York: ACM.

Hogg, M. A. (1996). Social identity, self-categorization, and the small group. In E. H. Witte & J. H. Davis (Eds.), *Small group processes and interpersonal relations* (Vol. 2, pp. 227–253). Hillsdale, NJ: Erlbaum.

Hogg, M. A., & Terry, D. J. (2000). Social identity and self-categorization processes in organizational contexts. *Academy of Management Review, 25*, 121–140.

Johnson, S. L., & Faraj, S. (2005). Preferential attachment and mutuality in electronic knowledge networks. *Proceedings of the 26th International Conference on Information Systems*. AIS.

Jones, Q., Ravid, G., & Rafaeli, S. (2002). An empirical exploration of mass interaction system dynamics: Individual information overload and usenet discourse. *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*. IEEE Computer Society.

Jones, Q., Ravid, G., & Rafaeli, S. (2004). Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research, 15*, 194–210.

Kalman, Y. M., Ravid, G., Raban, D.R., & Rafaeli, S. (2006). Pauses and response latencies: A chronemic analysis of asynchronous cmc. *Journal of Computer-Mediated Communication, 12*, 1–23.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality & Social Psychology, 65*, 681–706.

Kollock, P. (1999). The economies of online cooperation: Gifts and public goods in cyberspace. In M. A. Smith & P. Kollock (Eds.), *Communities in cyberspace* (pp. 259–262). London, UK: Routledge.

Kozlowski, K. J., & Klein, S. W. J. (2000). *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. San Franscisco, CA: Jossey-Bass.

Kraut, R. E., & Resnick, P. (2012). *Evidence-based social design: Mining the social sciences to build online communities*. Cambridge, MA: MIT Press.

Kraut, R. E., Sunder, S., Telang, R., & Morris, J. (2005). Pricing electronic mail to solve the problem of spam. *Human–Computer Interaction, 20*, 195–223.

Kraut, R. E., Wang, X., Butler, B. S., Joyce, E., & Burke, M. (2007). *Building commitment and contribution in online groups through interaction* (Working paper). Carnegie Mellon University, Pittsburgh, PA.

Lackaff, D. J. (2005). Norm maintenance in online communities: Analysis of heterarchical moderation regimes (Unpublished master's thesis). La Trobe University, Victoria, Australia.

Lakhani, K. R., & Von Hippel, E. (2003). How open source software works: "Free" user to user assistance. *Research Policy, 32*, 923–943.

Lampe, C., & Johnston, E. (2005). Follow the (slash) dot: Effects of feedback on new members in an online community. *Proceedings of ACM Conference on Supporting Group Work*. New York: ACM.

Lampe, C., Johnston, E., & Resnick, P. (2007). Follow the reader: Filtering comments on slashdot. *Proceedings of the CHI 2007 Conference on Human Factors in Computer Systems*. New York: ACM.

Liang, T. P., Lai, H. J., & Ku, Y.C. (2007). Personalized content recommendation and user satisfaction: Theoretical synthesis and empirical findings. *Journal of Management Information Systems, 23*(3), 45–70.

Ling, K., Beenen, G., Ludford, P. J., Wang, X., Chang, K., Li, X., . . . Kraut, R. E. (2005). Using social psychology to motivate contributions to online communities. *Journal of Computer Mediated Communication, 10*(4), article 10.

Lott, A. J., & Lott, B. E. (1965). Group cohesiveness as interpersonal attraction: A review of relationships with antecedent and consequent variables. *Psychological Bulletin, 64*, 259.

Maloney-Krichmar, D., & Preece, J. (2005). A multilevel analysis of sociability, usability, and community dynamics in an online health community. *ACM Transactions on Computer-Human Interaction, 12*, 201–232.

Markus, L. (1987). Towards a "critical mass" theory of interactive media: Universal access, interdependence, and diffusion. *Communication Research, 14*, 491–511.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283–308). New York, NY: Academic Press.

North, M. J., & Macal, C. M. (2007). *Managing business complexity: Discovering strategic solutions with agent-based modeling and simulation*. London, UK: Oxford University Press.

Pew Internet. (2001). Online communities: Networks that nurture long-distance relationships and local ties. Retrieved from http://www.pewinternet.org/PPF/r/47/report_display.asp

Postmes, T., & Spears, R. (2000). Refining the cognitive redefinition of the group: Deindividuation effects in common bond vs. common identity groups. In T. Postmes, R. Spears, M. Lea, & S. Reicher (Eds.), *SIDE Effects Centre Stage: Recent Developments in Studies of De-individuation in Groups* (pp. 63–78). Amsterdam, the Netherlands: KNAW.

Preece, J. (2000). *Online communities: Designing usability, supporting sociability*. Chichester, England: Wiley.

Preece, J., Nonnecke, B., & Andrews, D. (2004). The top 5 reasons for lurking: Improving community experiences for everyone. *Computers in Human Behavior, 20*, 201–223.

Prentice, D. A., Miller, D. T., & Lightdale, J. R. (1994). Asymmetries in attachments to groups and to their members: Distinguishing between common-identity and common-bond groups. *Personality & Social Psychology Bulletin, 20*, 484–493.

Raban, D. R., Moldovan, M., & Jones, Q. (2010). An empirical study of critical mass and online community survival. *Proceedings of the CSCW 2010 Conference on Computer Supported Cooperative Work*. New York: ACM.

Ren, Y., Kraut, R. E., & Kiesler, S. (2007). Applying common identity and bond theory to design of online communities. *Organization Studies, 28*, 377–408.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of netnews. *Proceedings of the CSCW 1994 Conference on Computer Supported Cooperative Work*. New York: ACM.

Resnick, P., & Zeckhauser, R. (2002). Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. *The Economics of the Internet and E-Commerce, 11*, 127–157.

Ridings, C. M., & Gefen, D. (2004). Virtual community attraction: Why people hang out online. *Journal of Computer-Mediated Communication*, 10(1), article 4.

Roberts, J., Hann, I., & Slaughter, S. (2006). Understanding the motivations, participation, and performance of open source software development: A longitudinal study of the Apache projects. *Management Science, 52*, 984–1000.

Rogers, E. M., & Agarwala-Rogers, R. (1975). Organizational communication. In G. L. Hanneman & W. J. McEwen (Eds.), *Communication behavior* (pp. 218–236). Reading, MA: Addison-Wesley.

Sassenberg, K. (2002). Common bond and common identity groups on the internet: Attachment and normative behavior in on-topic and off-topic chats. *Group Dynamics, 6*, 27–37.

Schneider, B., Goldstein, H., & Smith, D. (1995). The ASA framework: An update. *Personnel Psychology, 48*, 747–773.

Smith, M. A. (1999). Invisible crowds in cyberspace: Mapping the social structure of the Usenet. In M. A. Smith & P. Kollock (Eds.), *Communities in cyberspace* (p. 203). London, UK: Routledge.

Sterman, J. D. (2002). All models are wrong: Reflections on becoming a systems scientist. *Systems Dynamic Review, 18*, 501–531.

Turner, T. C., Smith, M. A., Fisher, D., & Welser, H. T. (2005). Picturing Usenet: Mapping computer-mediated collective action. *Journal of Computer-Mediated Communication, 10*(4), article 7.

Van Alstyne, M., & Brynjolfsson, E. (1996). Electronic communities: Global village or cyberbalkans? Retrieved from http://www.mit.edu/people/marshall/papers/CyberBalkans.pdf

Von Ahn, L., Blum, M., & Langford, J. (2004). Telling humans and computers apart automatically. *Communications of the ACM, 47*(2), 56–60.

Vroom, V., Porter, L., & Lawler, E. (2005). Expectancy theories. In J. B. Miner (Ed.), *Organizational Behavior 1: Essential Theories of Motivation and Leadership* (pp. 94–113). New York, NY: M.E. Sharpe.

Walther, J. B. (1994). Anticipated ongoing interaction versus channel effects on relational communication in computer-mediated interaction. *Human Communication Research, 204*, 473–501.

Wang, X., Butler, B. S., & Ren, Y. (2013). The impact of membership overlap on the growth: An ecological competition view of online groups. *Organization Science, 24*, 414–431.

Wang, Y.-C., & Kraut, R. E. (2012). Twitter and the development of an audience: Those who stay on topic thrive! *Proceedings of the CHI 2012 Conference on Human Factors in Computer Systems.* New York: ACM.

Wasko, M., & Faraj, S. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly, 29*, 35–58.

Wellman, B., & Gulia, M. (1999). Virtual communities as communities: Net surfers don't ride alone. In M. A. Smith & P. Kollock (Eds.), *Communities in cyberspace* (pp. 167–194). London, UK: Routledge.

Wilensky, U. (1999). NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Worthen, B. (2008). Why most online communities fail? Retrieved from http://blogs.wsj.com/biztech/2008/07/16/why-most-online-communities-fail/

Wright, K. (2000). The communication of social support within an on-line community for older adults: A qualitative analysis of the SeniorNet community. *Qualitative Research Reports in Communication, 1*, 33–43.