Under Review at *Information Systems Research*
Please do not cite without author permission

Agent-Based Modeling to Inform Online Community Theory and Design: Impact of Discussion Moderation on Member Commitment and Contribution

Yuqing Ren
Carlson School of Management
University of Minnesota
Minneapolis, MN 55455

Robert E. Kraut
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA 15213

ABSTRACT

In this article, we advocate a new approach in theory development by translating and synthesizing insights from multiple social science theories in an agent-based model to understand challenges in building online communities. To demonstrate the utility of this approach, we use it to examine the effects of three types of discussion moderation in conversation-based communities: no moderation, in which all members are exposed to all messages, community-level moderation, in which off-topic messages are deleted for everyone in the group, and personalized moderation, in which people see different messages based on their interests. Our results suggest that personalized moderation out-performs the others in increasing members' contribution and commitment, especially in topically broad communities and those with high message volume. In comparison, community-level moderation increases commitment but not contribution. Our results also reveal a critical trade-off between informational and relational benefits. This research demonstrates the value of agent-based modeling in synthesizing more narrowly-focused theories to describe and prescribe behaviors in a complex system, to generate novel theoretical insights that were out of scope for the component theories, and to use these insights to inform the design of online communities.

Key words: online community, agent-based modeling, design, motivation, simulation

1. Introduction

The Internet provides a new and popular platform for social interactions. According to a Pew Internet report (2001), 84% of American Internet users, or about 90 million adults participate in online communities to share information, exchange social support, and discuss hobbies, politics, sports, and other topics of interest. Following Preece (2000), we define an online community as an Internet-connected collective of people who interact over time around a shared purpose, interest, or need. Although many online communities are highly successful, many others fail. For example, across a wide range of Usenet groups, more than 60% of the newcomers who post a message in one month in a group never post again (Arguello et al. 2006). A recent Deloitte survey of more than 100 businesses attempting to build online communities, some spending over $1 million in the effort, found that most efforts failed to attract a critical mass of users. About 35% of the communities studied had fewer than 100 users, and only 25% had more than a 1,000 (Worthen 2007).

An important reason behind these failures is the lack of evidence-based, practical guidance in 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 the decisions and fail to anticipate how users will respond. For instance, if a community offers points for contributions and recognizes the biggest contributors on a public “leader board,” this feature may encourage average participants to increase their level of contribution to the level of others; on the other hand, it could discourage them from contributing if they perceive their efforts as unnecessary, since the leaders are providing sufficient content.

The two predictions originate from two social science theories: social comparison theory (Festinger 1954) and the Collective Effort Model (Karau and Willams 1993). The former argues that people have a tendency to evaluate their performance against others and are likely to increase their effort when being told that others have contributed more (Harper et al. 2007). The latter argues that people exert less effort in groups because they perceive their efforts are unnecessary to achieve group outcomes. While 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 solutions for building successful communities (see Ling et al. 2005 for an exception). An important reason, we suspect, may be that the logic of design, which manages trade-offs among tens or hundreds of parameters that can influence members' behaviors, 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 thus requires a navigation map that synthesizes insights from multiple theories to identify the pathways through which particular design choices affect the different outcomes that designers aim to achieve.

In this paper, we advocate a new approach in both theory development and community design by translating and synthesizing propositions from component social science theories into an agent-based model that can help understand challenges in building online communities. We demonstrate the usefulness of this new approach by applying the model to understand an important design decision – the type of moderation designers should impose in conversation-based online communities, such as a movie discussion newsgroup or social support web forum[1]. We examined three ways of moderating online discussion – no moderation, community-level moderation, and personalized moderation – and how they affect member commitment and community activity. Simulation results highlight the general superiority of personalized moderation over designs with no moderation or community-level moderation and provide practical guidelines in choosing moderation mechanisms to match community characteristics such as topical breadth and message volume. The results also reveal a critical trade-off between designing for informational benefits and designing for relational benefits.

[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.

This exercise makes four contributions to the information systems literature. First, synthesizing propositions from multiple component theories in an agent-based model produces a new multilevel theory of online communities, one that generates new theoretical insights for understanding and designing these communities. Second, we contribute a systematic understanding of how different styles of discussion moderation affect member commitment and community activity that can be applied to guide design practices. Even though we examined only conversation-based communities, we believe that the benefits of personalization, the contingencies of design decisions on community characteristics, and the trade-offs between informational and relational benefits will apply to other types of communities as well. Third, the agent-based model, based on a carefully selected set of component theories, serves as a test bed, which can be extended in future research to incorporate a broader set of theories and to inform a wider variety of design choices, such as how to socialize newcomers or whom and what to feature on a leader board. Finally, we hope that our study will call attention to agent-based modeling as a powerful tool in theory integration and development, but one that has been underutilized in the Information Systems discipline (see Rao et al. 1995, Raghu et al. 2004 for exceptions).

The rest of the article is organized as follows. In sections 1.1 and 1.2, we elaborate the rationale for theory-driven design by showing a gap between social theories and design practice and how agent-based models enable us to combine insights from multiple theories to understand design trade-offs. Section 2 describes the conceptual framework for the model, its theoretical basis, and how we implemented and calibrated it. Section 3 describes simulation experiments and our main results comparing three types of discussion moderation in communities with different characteristics. Section 4 discusses the theoretical contributions and practical implications of our findings and how agent-based modeling informs critical trade-offs in online community design.

1.1. Why Theory-Driven Design?

Scholars have used many social science theories to understand behavior in online communities. For example, Ren et al. (2007) and Sassenberg (2002) use theories of group identity and interpersonal bonds to examine the development of members' commitment to online communities. Kollock (1999) and Ling et al. (2005) use public goods economics and theories of social loafing to analyze problems of under-contribution. Others use signaling theory to understand how reputation influences community success (e.g., Friedman and Resnick 2001, Ma and Agarwal, 2007). The success of these analytic exercises demonstrates the applicability of traditional social science theory to the new phenomenon of online communities.

Even though social science theory is being used to describe existing communities, it is rarely used prescriptively, as the basis for designing them. A major reason is that the logic of design, which attempts to manage trade-offs among tens or hundreds of 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, holding everything else equal. This *ceteris paribus* paradigm for developing and testing social science theory 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 period.

Consider attempts to understand and remediate problems of under-contribution, which are endemic in many online groups (Ling et al. 2005). Contribution, like most outcomes of interest, has multiple causes, and each cause is treated by separate social science theories. Social psychologists who developed the collective effort model to explain social loafing, for example, concentrate on contributors' identifiability, uniqueness of contribution and liking for the group (Karau and Williams 1993). Public goods theories concentrate on expected utilities that cause people to contribute (Ledyar 1993). Contribution is also influenced by the psychological attachment of community members, the presence of explicit goals, group norms, intrinsic motivations and extrinsic incentives, among many other factors. Thus multiple theories are needed to model contribution in online communities and to design effective interventions to increase contribution.

On the other hand, a single design choice can have conflicting effects on contribution and other unforeseen outcomes. For example, the collective effort model proposes that people will contribute *more* to groups that they like. Therefore, increasing the homogeneity among group members may cause members to contribute more because homogeneity leads to liking and thus greater willingness to expend effort to help the group and its members. However, the collective effort model also proposes that people will contribute more when they believe their contributions are unique and thus necessary for group success. As a result, they may contribute *less* in a homogeneous group because they feel their efforts are redundant. Homogeneity may also influence commitment to the group and ease of communication, which have effects on important outcomes beyond contribution, including turnover rates or production efficiency. Theory-driven design requires a broad navigation map that synthesizes insights from multiple theories to identify the pathways through which particular design choices may have positive and negative effects on the different outcomes that designers aim to achieve.

1.2. Theory Synthesis in an Agent-Based Model

In this article, we present an agent-based model to express, synthesize, and extend social psychological theories that are relevant to motivation and contribution in online communities. Agent-based models capture the behaviors of complex adaptive systems from the ground up (North and Macal 2007). The emergent properties of a complex social system (e.g., a financial market, beehive or online community) are examined by simulating the behaviors of the agents that comprise the collective (e.g., the traders, bees or members). Compared with conventional methods, agent-based modeling is especially suitable for bottom-up theorizing (Kozlowski and Klein 2000) to understand how individual behaviors interact over time and lead to emergent system-level patterns, and to predict potential outcomes of these interactions.

The agent-based model described in this article simulates the behaviors of individual members of an online community in order to understand how various design interventions affect community success. Agents in the model are animated using principles derived from well-established social science theories. The collective effort model proposes that members contribute to groups to obtain valuable outcomes (Karau and Williams 1993). Theories of group identity and interpersonal bonds propose that members are likely to contribute to groups if they feel psychologically attached to the group or its members (Prentice et al. 1994). Information overload theory proposes that human beings' information processing capacity is limited and too many messages or noise in communication requires extra effort (Rogers and Agarwala-Rogers 1975). Synthesizing multiple theories enables us to trace the paths through which a design choice may affect members' motivation. The use of an agent-based model also opens up the black box between design choices and visible outcomes, by modeling intervening variables, such as the benefits that members receive when they participate in and contribute to online communities.

The agent-based model is a new theory of online communities that helps to understand the complex, reciprocal interdependencies between the behaviors of individual members and properties of the community as a whole as it develops and evolves over time. That is the model is simultaneously a theory of motivation and commitment of individual members in the community and a theory of community growth and sustainability. A design choice can have both immediate, first-order effects (e.g., identifying members increases their contributions, e.g., Williams et al. 1981) and longer-term, second-order effects (e.g., increased contribution leads to information overload and drives other members away, e.g., Jones et al. 2004). The agent-based model can also serve as a test bed (Rao et al. 1995) for running what-if experiments, to inform the design of online communities. Integrating existing theories in an agent-based model enables one to identify places where the component theories agree, disagree, or are independent of each other, and to pin down factors that community designers could manipulate to produce desirable outcomes.

1.3. Discussion Moderation as a Design Decision

To demonstrate the usefulness of agent-based modeling in theorizing about and designing online communities, we apply it to understand the effects of different types of discussion moderation on member commitment and community activity. At the core of most online communities are members who converse to ask and answer questions, exchange opinions and social support, to develop personal relationships and to enjoy each others' company. Without conversation, these communities would vanish. Even in online games like World of Warcraft or production-oriented communities like Wikipedia, members depend upon conversations to coordinate their work and to develop commitment to the group.

Even though 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 heterogeneity of topics can drive people away from online communities (Butler 2001, Jones et al. 2004). High message volume can be especially problematic when much of the conversation is off-topic or when the community is organized to cover many different topics. Although many communities are organized around specific topics, people in them often engage in personal, off-topic conversations that have nothing to do with the nominal topic. For example, a large minority of messages in a discussion group organized around depression may have nothing to do with depression (http://discussions.seniornet.org, Wright 2000), and messages in investment discussion groups may have little to do with finances (Gu et al. 2007). For members interested only in the nominal topic of the community, off-topic messages are an irritation that can drive them away. High message volume is also problematic in communities that encourage conversation across a wide range of topics. As Butler (1999) notes, in communities treating many topics, messages interesting to some members are likely to be off-topic and uninteresting to others.

The default view in most online discussion communities shows every member all messages, organized either in chronological order or in threads. To deal with problems of high message volume and off-topic conversation, designers and managers of online communities have augmented the default with techniques for moderating the discussion. A common practice is community-level moderation, in which case human moderators or software agents block or remove inappropriate or off-topic messages (Figallo 1998, Lampe and Johnston 2005). A message is available either for everyone visiting the site or for 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 messages or other inappropriate material, such as spam, trolling messages or anti-social flames. Community-level moderation can be less effective in communities that attract members with diverse interests or offer diverse content. In such broadly defined communities, nominally on-topic messages might be of no interest to a large proportion of members. For example, in the movie discussion forum rottentomatoes.com, a message evaluating a new action movie is likely to be of no interest to members who dislike action movies. Conversely, nominally off-topic conversations, such as one describing high school romances consummated in movie theatres, may be of great interest to some members. Under either scenario, community-level moderation leads to sub-optimal user experience.

A less common yet promising practice is personalized moderation, in which case different users get to view different subsets of messages matched to their interests. In e-commerce sites, personalized recommendation increases users' satisfaction by decreasing the total number of items to be processed and thus reducing information overload, while at the same time increasing each item's average fit to users' interests (Tam and Ho 2005, Liang et al. 2007, Schafer et al. 2001). We believe personalized moderation may have the same effect on user satisfaction and motivation in the context of online discussion groups. For instance, software agents can be created to automatically match messages against a static personal profile or one that is dynamically updated based on user behaviors in the community[2] (e.g., Harper et al. 2007).

In this article, we examine discussion moderation as a way to demonstrate the benefits of using an agent-based model as a theory of online communities, one that can be used to provide practical guidance in designing online communities. By representing an online community as an agent-based model synthesizing component social science theories, we aim to answer three ques-

[2] We examine personalized moderation at the conceptual level. Its implementation is beyond the scope of this paper. Other ways to implement may include collaborative filtering, as used by digg.com or slashdot.com, where members rate messages so that others can use these ratings to guide their reading, or participation structure, as used by rotttentomatoes.com, where members and messages are segmented into sub-forums.

tions. (1) How does the style of moderation affect a community's viability and its members' commitment? (2) To what extent are the effects of moderation contingent upon community characteristics such as topical breadth and message volume? (3) How does the style of moderation affect the trade-offs among the various benefits that members receive from participating in an online community? Below we describe the conceptual framework of the model and its theoretical basis.

2. The Conceptual Framework for the Agent-Based Model

Figure 1 depicts the conceptual framework underlying the agent-based model of motivation in online communities. In this model, the benefits that members think they will receive from participating in the community influence their motivation to return to the community and to read and post messages. These individual decisions result in an online community with certain characteristics at a particular time, such as the numbers and quality of messages in it, the number of members in it and their activity, and the similarity of among members and the frequency of their interactions. Design interventions, such as ways in which the community moderates discussion, also influence these community characteristics. These emergent features of the community in turn influence the benefits participants think they will receive in the future.

This model is based on several social psychological theories of motivation and voluntary contribution. We started with the most basic motivation theory, expectancy-value theory (Vroom, 2005), and used one of its extensions, the Collective Effort Model (Karau and Williams 1993), which addresses how contribution to a group influences both the net value participants receive from their contributions and their expectation of this value. The Collective Effort Model lays the basis for the theoretical framework by suggesting that people contribute to a group to the extent that they believe their efforts will directly or indirectly lead to benefits to themselves. Yet the Collective Effort Model is silent about the different types of benefits members derive from participating in a group. Empirical research (Ridings and Gefen 2004) suggests three types of benefit people derive from participating in online communities: (1) informational benefits from accessing and sharing information with others; (2) social benefits from identifying with a group and interacting with its members; and (3) other benefits from recreation and reputation. The Collective Effort Model does not differentiate liking of the group from liking its members. We therefore draw insights from group identity and interpersonal bonds theories to calculate social benefits, and insights from information overload theory to calculate informational benefits. After benefits are calculated, the model calculates member motivation as a weighted sum of the benefits, with the weights indicating how much that member values each type of benefit.

Insert Figure 1 about Here

Due to the complexity of the model, we describe it in three steps. We first describe the decision rules that determine whether an agent reads or posts a message in the community. We then describe the calculation of the benefits the agent receives from membership, which determine motivation to read and post messages. We finally describe the model implementation and simulation experiments. For convenience, we describe how the model operates for a movie discussion forum. The model applies broadly, however, to text-based, conversationally-oriented online communities. In the description of the model below, we use the term “member” when describing people in an online community the model simulates and the term “agent” when describing decision rules implemented in the model.

2.1. Member Actions: Reading and Posting Messages

Table 1 provides an overview of the decision rules that an agent uses to decide whether to take various actions. Following Butler (2001), we define participation as an action that members take to be exposed to communication activity, such as reading messages. We define contribution as an action that members take to engage actively in community activity, such as posting messages. Following the utility-like logic underlying the collective effort model, we assume that a member (1) logs in to read messages when expected benefit from participation exceeds expected cost, and (2) posts messages when expected benefit from contribution exceeds expected cost.

Insert Table 1 about Here

2.1.1. Which messages to read? Typically, a member views messages in reverse chronological orders and stops viewing when he runs out of time, interest, or messages. We assume that the total number of messages a member views on a particular day depends upon the total number of messages available that day and how motivated the member is to read. We calculate messages an agent will view on a specific day as proportional to the amount of benefit he has received in the past from reading messages, capped by the total number of messages available to read. Because most people read in reverse chronological order and messages get stale with time, 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 et al. 2006, Kalman et al. 2006) . To post a message, an agent makes three additional decisions: (1) whether to start a new thread or reply to an existing post; (2) the topic of the message, if starting a new thread and, (3) which message to respond to, replying. For simplicity, we assume that the agent is equally likely to start a new thread or to reply to an existing thread, and sensitivity analyses indicate that our results remain robust when we vary the percentage of starting a new thread from 30% to 70%.

2.1.2. 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. A member can be interested in one or more of the topics. We assume that members' interests remain static and do not change during the experimental period. We assume that each message concerns only one topic, although the analysis is the same if each message refers to several topics. When an agent posts a thread-starting message, the topic of this message is a joint function of the agent's interests and the topics of messages the agent has recently viewed, to account for social influence. 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 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.

2.1.3. Which message to reply to? Theory and empirical evidence (Johnson and Faraj 2005, Fisher et al. 2006) suggest three common patterns of interaction among community members: (1) preferential attachment, in which members respond to popular messages or posters, (2) reciprocity, in which members respond to those who have written to them in the past, and (3) interest matching, in which members respond to messages that match their interests. Of course, people respond only to messages they have read. The agent in the model chooses to reply to a message based on a weighted sum of (1) the number of replies the message has received; (2) the number of times the poster of the message has responded to the agent; and (3) the match between the topic of the message and the agent's interests.

2.2. Member Benefits and Costs

Table 2 provides an overview of how informational, social, and other benefits are implemented in the model, including the theories used to make assumptions, the rules used to calculate different types of benefits, and key parameters in the benefit functions.

Insert Table 2 about Here

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

Benefit from accessing information. We assume that (1) only messages that match members' interests provide information benefit, and (2) benefit from accessing information is a marginally decreasing function of the number of messages viewed. According to information overload theory, human beings 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). We calculate the benefit from reading messages as a joint function of the quantity and quality of messages that an agent reads. On average, the more messages an agent reads that match his interest, the greater information benefit he receives, with diminishing returns, because of information redundancy or information overload. The first graph in Table 2 illustrates the information access benefit function. The parameters were based on Liang et al.'s (2007) experimental study of 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 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 thus 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, that is, 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 et al. 2006) or performing community maintenance tasks such as promoting and policing the site (Butler et al. 2007). Engaging in these altruistic actions can lead to positive self-evaluation of competence and social acceptance because it feels good to help others and the community (Wasko and Faraj 2005). It also provides opportunities to express one's values related to altruistic concerns for others and to exercise one's knowledge and skills that might otherwise go unpracticed (Clary et al. 1998). According to the collective effort model (Karau and Williams 1993), intrinsic benefits from contributing to group outcomes decrease if members believe that (1) the group is large or (2) others are already contributing, both of which make their help less necessary. On the other hand, intrinsic benefits from sharing information to help others and the community increase when people perceive group tasks as interesting, or when they identify with the group or like other members.

The pseudo code in Table 3 shows how we implement these rules in the model. If the agent is interested in the messages or feels strongly attached to the group or its members, we calculate two components – one is non-zero when the group is perceived as at risk of failing (operationalized as hosting fewer than 100 messages), and the other is non-zero when others are perceived as under-contributing. We assume that agents who have a history of contributing ten times more than the community average tend to perceive others as under-contributing and therefore compensate for others' lack of contribution. Finally, in order to capture the diffusion of responsibility effect, we divide the sum of all components by a marginally decreasing function of group size or the total number of others who are present to contribute.

Insert Table 3 about Here

2.2.2. Benefit 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 members to become attached to groups and contribute toward group efforts (e.g., Prentice et al. 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 2000). Shared interests and similarity in preferences have been used to manipulate and measure identity in laboratory experiments (Amichai-Hamburger 2005, Postmes and Spears 2000). To simplify the model, we assume that people who share a common interest with the community identify with it. For example, a movie lover feels a stronger sense of belonging to a discussion group if other members are also movie lovers and if the conversation is about their shared interests than if the forum is full of discussion of jobs, love, politics, or other off-topic subjects. 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 level of identity-based attachment the agent feels to the community.

Benefit from bond-based attachment. Research on small groups suggests that repeated interaction leads to interpersonal attraction (Festinger et al. 1950). As the frequency of interaction between two persons increases, their liking for one another also increases (Cartwright and Zander 1953). Studies of Usenet groups suggest that getting a quick reply after posting seems to encourage members of an online community, especially newcomers, to return and participate in community discussion (Kraut et al. 2007). We speculate that replies from other members signal the likelihood of forming relationships with others in a community. We thus calculate the benefit an agent receives from interpersonal bonds as a function of the number of other agents with whom the agent has developed a relationship through repeated, mutual interaction (i.e., the two agents have responded to each other at least twice) and the number of responses the agent has received during the last period of interaction, whichever is higher. Because both attitude similarity (Byrne, 1997) and personal self-disclosure (Collins & Miller, 1994) lead to liking, the two interacting agents will like each other more if they have similar interests or if the interaction involves off-topic subjects. Benefit from interpersonal bonds takes a marginally decreasing form, as illustrated in Table 2: The first few relationships an agent develops bring greater social benefit than subsequent ones.

2.2.3. Benefit from recreation. A third motivation that leads people to join online communities is recreational, that is, the enjoyment members derive from reading and posting online (Ridings and Gefen 2004). Several studies have identified stable individual differences in the extent to which people think online behavior is fun (e.g., Cotte et al. 2006). For instance, posters enjoy online interaction more than lurkers do (Preece et al. 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 Table 2). With a gamma distribution, the majority of members have a moderate level of interest in reading and posting messages in online communities; and a small proportion of members have a high level of interest.

2.2.4. Benefit from reputation. People are also motivated to contribute to online communities by the reputation they gain by doing so (Wasko & Faraj, 2005). 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 are among the top ten percent of contributors receive reputation benefit. Sensitivity analyses indicate that the main results were robust when the proportion receiving reputation benefit varied between 5% and 15%.

2.3. Costs, Motivation, Member Entry and Exit

2.3.1. Motivation as a weighted sum of benefits. As mentioned earlier, agents' motivation to read and to post is calculated as a weighted sum of their benefits from reading and posting minus the costs of reading and posting. These weights differ across communities (Ridings and Gefen 2004). In the model, we set the weights for information exchange, identity, bonds, and recreational and reputational benefits at 0.5, 0.1, 0.3, and 0.1 respectively; these weights are consistent with Riding and Gefen's (2004) finding about interest communities[3]. Within a single community, members have various reasons for joining. Some people may go to a movie discussion site for information about which movies to watch, others for dates and companionship, and yet others because they identify as a movie buff. In the model, weights for individual agents were drawn from normal distributions around the community means.

2.3.2. Costs of participation and contribution. We model three types of cost associated with reading and posting messages. Access cost simulates the time and effort people spend logging

[3]Note that the current model describes behaviors within an interest community, like a movie discussion group. We do not vary community type in the current research. To do 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 might be 0.5, 0.25, 0.1 and 0.15. In contrast, the weights for a cancer support group, where one's disease helps defines one's identity, .33. .33, .33 and 0.1, respectively.

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 work, conversation 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 mid-career wage earners than for teens or retires).

2.3.3. 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 some parameters relevant 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 an average or a smaller number of newcomers on most days, and attracts many newcomers on a small number of days. In our model, agents do not make conscious decisions to leave the community. If their experienced benefits are below a threshold, they simply stop coming back.

2.4. Model Implementation and Calibration

We implemented the simulation using NetLogo, a cross-platform multi-agent modeling environment (Wilensky 1999). Within the simulation, agents take actions during a simulated day, following the sequences depicted in the Appendix. All active agents in the simulated community are given the opportunity to make a reading and posting decision before anyone moves to the next day. Messages posted the previous day are distributed to all agents the next day and 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 insure 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 described 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. We describe this calibration in more detail below. To assure the robustness of our results, we ran a series of sensitivity analyses by relaxing key assumptions and varying key parameters. Results do not differ substantially from those we report in the results section.

2.4.1. Model calibration and validation. 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). Previous studies show that three statistics describing online communities – posts per member, replies per post, and communication partners (out-degrees) per member – demonstrate a power-law distribution (Fisher et al. 2006, Smith 1999). We use these three stylized facts to calibrate the model. Our calibration and validation process involved tweaking parameters in the model so that it generated simulated data that matched training data from 12 Usenet groups and then testing 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 twelve online groups starting with data on membership size and message volume from twelve Usenet groups. The groups had 30 to 500 posters and 30 to 500 active messages in the beginning of the simulated period. We then engaged in an iterative process in which we compared the distribution of the three statistics – posts per member, replies per post, and out-degrees per member – from the simulation 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 our intuitions. After ten 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 and Obel 1995), which greatly increases the construct and external validity of our model.

We then simulated another 25 online groups, starting with data on membership size and message volume from real groups. The simulated statistics fit the real statistics for these 25 groups well and demonstrate the validity of the model. Figure 2 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 2, for all three statistics. The coefficients range between 0.9 and 0.96, confirming a good match between the empirical data and simulated data. We also examined survival curves for members and messages during model calibration and validation and found similarity between simulation and real data as well. As shown in Figure 3, the survival curve from real data suggests that about 60% of new posters fail to return after their first post, and on average about 10% to 20% posters stick around for over 100 days

Insert Figure 2 and Figure 3 about Here

3. Simulation Experiments and Results

3.1. Virtual Experimental Design

In this section, we describe a full-factorial simulation experiment examining the effects of discussion moderation on community performance when topical breadth and message volume vary. We simulated three levels of topical breadth, in one, five, or nine topics were germane to each group respectively, and three levels of message volume with on average about 10, 15, and 20 messages per day. We simulated three types of moderation: no moderation, community-level moderation (under which messages whose topics do not conform to community purpose are removed), and personalized moderation (under which a personalization algorithm presents a subset of messages that match a member's interests).

Here is an example to illustrate the three ways of moderating online discussion. Imagine a movie discussion forum with five germane topics: upcoming movies, upcoming DVDs, movie critics, celebrities, and video games. Tony is an active member interested in upcoming movies and celebrity gossips. In a group with no moderation, Tony sees all messages regardless of topics, in reversed chronological order with the latest being on top. With a group using community-level moderation, Tony see only on-topic messages because an administrator or moderator reviewed messages and removed those that do not address the legitimate topics (e.g. rants about publishers' copy protection policies). Filtering is at the community level, and all other community members see what Tony sees. Under personalized moderation, Tony sees only messages having to do with upcoming movies and celebrities, and can also choose to be exposed to stories about copy protection if he finds them interesting. Filtering occurs at the individual level, and other members can chose a different selection of messages.

We ran a 365-day simulation for each experimental condition on five[4] randomly constructed groups. All groups began with 30 seed members and 30 seed messages and evolved over time as newcomers joined and old-timers left. At each simulated day, each agent assessed prior benefits from having read and posted messages and decided whether to login to read and post messages. For purposes of the simulation, the precision of the personalized moderation was set to 80% of recommended messages matching a member's interests. Sensitivity analyses suggest that the main results remain robust when the precision of personalized moderation varies between 60% and 100%.

We examined the effects of conversation moderation on two outcomes that would easily be visible to any community manager: the number of new posts per day, which is an indictor of com-

[4] We chose five as a convenient yet arbitrary number. So we did two things to make sure that our results are substantial and not a result of the number of groups we simulate. First, we examined both the level of significance and effect sizes that are independent of sample sizes (Cohen 1988). All effects that are significant at $p < 0.05$ level have an effect size greater than 0.5 standard deviation. Second, we ran a separate experiment with 20 groups per conditions. All results remained unchanged.

munity activity, and the average number of login sessions per member, which is an indicator of member commitment. We also examined the benefits members received from being part of the community, often invisible to community managers. We ran two two-way ANOVA analyses to examine the main effects of moderation and its interactions with topical breadth and message volume. 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 information access benefits and bonds benefits that our simulation results revealed as mediators, to understand the link between experimental conditions and community outcome measures.

3.2. Results

3.2.1. Effects of Moderation on Community Activity. Analyses of the impact of moderation on *community activity* (posts per day) in communities differing in *topical breadth* revealed a significant main effect of moderation. As shown in Figure 4, personalized moderation led to the highest level of community activity (15 to 20 posts per day), about 50% more than community-level moderation and 36% more than no moderation ($p < .001$). There was no significant difference between community-level moderation and no moderation ($p = .24$). The analyses also revealed a significant interaction between moderation and topical breadth ($p = .05$). Personalized moderation led to 67% more posts than community-level moderation and no moderation in communities with moderate (five topics) and high topical breadth (nine topics), compared to 0% to 25% more posts in communities with a single nominal topic. Topical breadth by itself had no effect on community activity ($p = .65$).

Insert Figure 4 about Here

Analyses of the impact of moderation on community activity in communities differing in *message volume* revealed two main effects and a significant interaction between moderation and message volume, as shown in Figure 5. By definition, communities with high message volume had more posts per day than communities with low volume (20 messages versus 10 messages, $p < .001$). Personalized moderation, again, led to more posts than community-level moderation and no moderation ($p < .001$). The interaction effect indicates that the difference was much greater in communities with higher message volume ($p < .001$). Personalized moderation led to 44% to 63% more posts than community-level moderation and no moderation in communities with high message volume, compared to 22% more posts in communities with low message volume.

Insert Figure 5 about Here

3.2.2. Effects of Moderation on Member Commitment. Analyses of the impact of moderation on *member commitment* (login sessions) in communities differing in *topical breadth* revealed a significant main effect of moderation ($p < .001$) and a significant interaction between moderation and topical breadth ($p < .001$). As shown in Figure 6, both personalized and community-level moderation led to more login sessions than no moderation, but under different conditions. Community-level moderation led to the highest login frequency in communities with a single topic (18% higher than no moderation) whereas personalized moderation led to the highest login frequency communities with more topics (32% higher than no moderation). Both differences were statistically significant at the $p < .001$ level. Topical breadth had no significant effect on member commitment ($p = .45$).

Insert Figure 6 about Here

Analyses of the impact of moderation on commitment in communities differing in *message volume* revealed two main effects and a significant interaction between moderation and message volume ($p < .001$). Higher message volume led to more frequent logins – from an average of approximately six logins in communities with low message volume to eight logins in communities with high message volume. Personalized moderation led to more frequent logins than community-level moderation, and community-level moderation led to more frequent logins than no moderation. Both differences are significant at the $p < .05$ level. The effect of personalized moderation, as shown in Figure 7, was greater in communities with medium and higher message volumes than in communities with low message volume ($p = .01$). Compared to no moderation, personalized moderation increased login frequency by 33% in communities with high message volume and by 14% in communities with low volume.

Insert Figure 7 about Here

3.3. Member Benefits from Information and Interpersonal Bonds

Posts and logins are observable behaviors. To better understand the route by which moderation affects posting and login behaviors, we examined their impact on two benefits – benefit from accessing information (informational benefit) and benefit from interpersonal bonds (relational benefit). Doing so illustrates an important design trade-off involved in choosing among moderation techniques.

3.3.1. Member benefit in communities with different topical breadth[5]. Figure 8 shows the effects of moderation and topical breadth on the amount of informational and relational benefit that agents received, averaged across all active members and the five snapshots at which benefits were recorded.

Agents received greater *informational benefit* in the topically broad communities than the topically narrow ones ($p < .001$), and with either type of moderation than with no moderation ($p < .001$). The interaction between moderation and topical breadth indicates community-level moderation and personalized moderation led to higher informational benefits under different conditions ($p < .001$). 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 10-15% more benefit compared to community-level or no moderation in communities with a broad focus.

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 ($p < .001$). Compared to community-level moderation, personalized moderation led to twice as much relational benefit, and no moderation led to approximately 75% more relational benefit. The effects of moderation on relational benefit also depended on different topical breadth ($p < .001$). Personalized moderation led

[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 8 and medium message volume in Figure 9 to make the figures more readable.

to 17% lower relational benefit than no moderation in communities with a narrow focus, but 43% more in communities with a broad focus.

Insert Figure 8 about Here

3.3.2. Member benefits in communities with different message volume. Figure 9 shows the effects of moderation and message volume on informational and relational benefit, averaged across active agents and snapshots.

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 while 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 did not vary substantially in communities with different levels of message volume.

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 ($p < .001$). The most striking result is the interaction between moderation and message volume ($p < .01$). As message volume increases from low to high, the effects of no moderation and community-level moderation remained roughly the same, whereas the positive effects of personalized moderation doubled, from being comparable or lower to the other two to 40% to 150% higher.

Insert Figure 9 about Here

4. Discussion

In this study, we synthesized several component social science theories in an agent-based model to express a new theory of motivation, contribution and commitment in online communities and used this theory to understand trade-offs in online community design. By building upon and integrating propositions from multiple theories, our model depicts a more complete picture than any single theory can depict on its own of how individual motivation and interactions affect community dynamics. Our effort was successful in at least three regards. First, our application of the model to understand how moderation affects community activity and member commitment led to plausible yet non-obvious predictions about the effectiveness of community-level and personalized moderation. Second, the availability of intermediate variables enabled us to examine not only the end results of design interventions but also how the results were produced, illustrating the critical trade-off between designing for informational benefit and designing for relational benefit. Third, in addition to providing a platform for examining different styles of discussion moderation, the model itself serves as a mid-level theory that can be further extended and applied to examine other design decisions such as leader boards, community size, and tactics for newcomer socialization, among others. Validity checks demonstrate the potential value of our model as a test bed to inform and assist online community design.

4.1. Contributions to the Online Community Literature

Table 4 summarizes the main findings. The first point worth noting is the largely positive effects of personalized moderation, which suggests that personalization is an under-exploited yet promising mechanism for managing online discussion. The positive effects were especially prominent in topically broad and high-traffic communities when members were at risk of information overload, from either an increased ratio of noise to useful information or too many messages. In comparison, community-level moderation was less effective than either its common use or experts' opinions would suggest. It did not encourage members to contribute more, and its positive effect on member commitment occurred only in narrowly defined groups.

Insert Table 4 about Here

The performance gap between personalized and community-level moderation can be partially explained by the trade-off between informational and relational benefits. As the analysis of benefits showed, one drawback of community-level moderation is that by removing off-topic messages, it increased the benefits that members received from information exchanges but reduced benefits they received from developing online relationships. This result is consistent with research in social psychology showing the importance of personal self-disclosure as both a cause and a consequence of personal relationships (Collins, 1994). In comparison, personalized moderation resolved the trade-off between exchanging information and establishing relationships by customizing experiences to match member's interests. None of the theories on which the agent-based model rests by itself would have predicted such effects.

The Collective Effort Model, for instance, would not have predicted the effects because it treats group variables (e.g., how much a person likes the group or its tasks) as exogenous and given rather than endogenous and driven by their experiences in the group. This limitation of the Collective Effort Model is in turn a limitation of the experimental laboratory approach on which it rests. Laboratory studies examine ad hoc, static groups that form in the laboratory and exist for only a short time period. Consequently, researchers make theoretical assumptions about certain variables acting as causes and others as consequences, even though in reality "assumed consequences" such as members' contribution to a group may feed back and drive "assumed causes" such as liking of the group and its tasks.

The insights we draw from theories of group identity, interpersonal bonds, and information overload complement the Collective Effort Model, although none of them alone would have predicted the benefits groups can derive from personalized moderation or the processes through which these benefits are produced. Theories about group identity and interpersonal bonds focus on the social benefits that members derive from their membership and would most likely recommend that online communities limit their size, recruit people who are similar to each other and impose no moderation of communication. The downside of this solution is that a narrowly focused community may not attract a critical mass of users or information to become sustainable. Information overload theory, on the other hand, focuses on the informational benefits that members derive from their membership and would most likely prescribe community-level moderation.

Our work applies and extends the component theories in the sense that we link these theories by treating the output variables in some theories as the input variables of another (e.g., similarity causes liking and liking causes contribution to group), and we examine multiple routes through which one input variable affects an output variables (e.g., similarity negatively influences motivation to contribute by making people believe that their contributions are redundant and positively influences it by strengthening identity-based or bond-based attachment to the community). The resulting theory is richer than the components theories because it links and synthesizes them through shared variables, and these interdependencies lead to predictions that none of the component theories would make by itself.

For similar reasons, none of the component theories alone would have predicted the interactions between moderation styles and community characteristics such as topical breadth and message volume. As in the case with organizational design more generally (Galbraith 1973), online community design involves many contingencies. There is no single best way to design and manage an online community. Rather, our findings show that each of the three moderation styles can be a good choice, depending upon community characteristics (topical breadth and message volume) and specific goals that designers wish to accomplish (to make members loyal or to increase their contribution). We speculate that the contingency effects are related to the trade-off between informational and relational benefits. Personalized moderation had stronger effects in communities with broad interests or heavy traffic because it was effective at promoting social benefit without reducing informational benefit, and vice versa. A high message volume, especially when covering a broad range of topics, can easily overwhelm members and make it hard for them to feel part of the community or connect with others with similar interests. Personalized moderation, by selectively showing members messages that match their interests, increases the likelihood of finding and interacting with similar others while decreasing the costs of having to filtering through a large number of messages to find the interesting ones.

The simulation produced two other unexpected results that need further investigation. The first is that community-level moderation led to greater commitment but not contribution. One possible reason is that community-level moderation may have led to differential retention of posters versus lurkers. Due to the public goods nature of online conversations, posters and lurkers have equal access to the 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. It is also possible that community-level moderation encourages on-topic posts yet discourages off-topic posts. If these two effects cancel each other out, there will be no net increase in posts. However, by removing off-topic messages, community-level moderation increases the signal-to-noise ratio and thus improves the overall experience of members who are interested in accessing valuable information. As a result, these members become more committed and visit the community more frequently.

The simulation also revealed an unexpected, negative effect of topical breadth on relational benefit. Simulation results suggest that in communities dealing with many topics, members seem to be less likely to engage in repeated interactions to form strong relationships, whereas in communities with a single topic, members' shared interest in the common topic can serve as a powerful bonding mechanism that leads to interpersonal liking. This effect is likely to be driven by interpersonal similarity. Members in narrowly defined communities will share more interests and thus like each other more than will those in broadly defined communities. This effect highlights the challenge in managing communities, such as health support groups, that have multiple goals (i.e., to distribute information and foster empathic relationships) and have a heterogeneous clientele with a diversity of motivations. In these communities, personalized moderation may help members to find and interact with other who share similar interests.

4.2. Implications on Online Community Design Practice

Our simulation results shed light on several challenges in designing and managing online communities. The first implication is that community-level moderation is less effective than either its 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 simulation experiments suggest that community-level moderation is effective only in narrowly defined communities. Community-level moderation, which uses the same rule for everyone in removing off-topic messages, improved the informational benefits agents received but lowered their likelihood of forming interpersonal bonds and thus of receiving relational benefit. In other words, community-level moderation promotes informational benefit at the expense of relational benefit.

In contrast, personalized moderation improves both informational and relational benefit, especially in communities that involve many topics and heavy traffic. It is a technology to handle the informational versus relational trade-off and can be effective even if the algorithms to predict interests are only moderately accurate (e.g., 60 percent precision in the simulation). The simulation results suggest that community managers can use personalized recommendations to create implicit clusters of users with similar interests so that members can choose to be exposed to a narrow or wide range of information, depending on their preferences, and return to their intimate circles for personal conversations. An increasing number of online communities are adopting this approach. For instance, newsvine.com[6] is a community in which members can rate and comment on news items from big and little media all around the world and to create their own columns as well. One of its offerings allows members to organize around common interests into Newsvine groups so that the site becomes a smaller and more personal place for them. It also provides four venues through which members can personalize their social circles with whom to share their new stories: the public, a public group, a private group, or a private circle of friends.

Another implication is that contingent and adaptive approaches are necessary to improve community design. There is no universally optimal design for all communities. 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, attract more members, and accommodate more topics, community designers and managers should consider personalized filtering to reduce information overload and assist members to find similar others to engage in interesting conversations. Even though personalized recommendations have been widely used to provide people with best personalized views of digital content, as in Google news and Amazon.com recommendations, it has been much underused to recommend online conversations.

[6] http://www.newsvine.com/_cms/help/groups

Personalization can be implemented in explicitly or implicitly. The explicit approach analyzes user interests using keywords and ratings they explicitly report. The challenge is that community members may not be willing or able to specify their interests. Many studies have documented the difficulties people have formulating profiles that retrieve just the content that they want (e.g., Foltz & Dumais, 1992). A survey of Internet users' attitudes toward personalization found that even though 81 percent of the respondents favored personalized content, only about 50% were willing to share personal or preference data (Greenspan 2004). Alternative approaches should be considered when user preference data become unavailable or too costly to obtain. For example, one might organize a heterogeneous community into subcommunities or organize sub-topics or use collaborative filtering to leverage others' opinions "to help individuals in the community more effectively identify content of interest from potentially overwhelming set of choices" (Herlocker et al. 2004).

Implicit approaches to personalization are another alternative. They analyze users' interests by mining patterns of usage, such as which messages they have read. These methods, combined with modern information retrieval techniques (e.g., Foltz & Dumais, 1992) and with collaborative filtering techniques that leverage similarities among community members (e.g., Konstan et al., 1997), are more accurate than simple keyword matches and do not require explicit effort on the user's part. Despite their potential, these implicit approaches may be under-used because they are more difficult to implement than simple restructuring of the community or the use of keyword-matching approaches.

4.3. Limitations

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

Although the model presented here was based on a selection of social science theories relevant to motivation, contribution and commitment in online communities, it was not exhaustive and did not incorporate all relevant theory. It was primarily based on social psychological theory, but the social science literature offers a wide range of theories on motivation, collective action, and group behaviors that could be exploited for community design. For example, theories of intrinsic motivation focusing on fun or flow could be incorporated into the current model (e.g., Csikszentmihalyi, 1997). Indeed, by including benefits from recreation, the current model includes a place holder which could be expanded to include principles from the literature on intrinsic motivation. Future research should extend the model to incorporate insights from other theories such as goal-setting, intrinsic and extrinsic rewards, and social comparison, deviant behaviors, norms and conformity, just to name a few. Even in the social psychological domain, we did not drill down to incorporate all that is known about a phenomenon. For example, the model treats relational benefits as an important motivator for participation in online communities. It models the formation of relationships as a function of similarity between two people and their frequency of interaction, but does not exploit much else that is known about relationship formation, such as the effects on liking of physical attractiveness, self-presentation, reciprocity, or balance.

We simulated one type of community: conversation-based ones 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 are exogenous and static. In real communities, however, newcomers who originally join to talk about the nominal topic of the group may increasingly value friendship with other community members after repeated encounters. Likewise, member interests or attitudes towards certain topics may shift over time in response to the messages to which they are exposed. In developing the simulation, we relied upon our best judgment to estimate key parameters in the benefit functions or distributions, when theory or empirical evidence was absent (Sterman, 2002). We ran a series of sensitivity analyses to assure that our main findings are robust and not dependent upon these parameters taking certain values. Our main results, including the positive effects of personalized moderation, remained unchanged, which further enhanced our confidence in the validity and reliability of the model. We mentioned some of the parameters that we varied in the sensitivity analyses in model description, and a complete list is available upon request.

4.4. 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 about designing and managing these communities are not motivated by a systematic understanding of member motivation and contribution but driven by intuition and trial and error, many communities are 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 and environments. In other words, we believe that online community design can go beyond intuition and trial and error and can benefit from the prescriptive power of social science theory. We believe that our model, incorporating multiple theoretical perspectives, has the potential to evolve into a multi-contingency tool for diagnosis and design of online communities (Burton and Obel 2004). 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. We hope, however, that a model of the type we describe in this article can serve as a test bed that can help designers gain some preliminary knowledge of which features to experiment with. 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 the social sciences for design.

References

Amichai-Hamburger, Y. 2005. Internet minimal group paradigm. *CyberPsychology & Behavior, 8(2,)*140-142.

Arguello, M., Butler, B., Joyce, E., Kraut, R., Ling, K.S., & Wang, X. 2006. Talk to me: Foundations for successful individual-group interactions in online communities. *ACM Conference on Human-Factors in Computing Systems*, New York: ACM Press. 959-968.

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(1,)* 57-71.

Burton, R.M., Obel, B. 2004. *Strategic organizational diagnosis and design: The dynamics of fit, with OrgCon software*. Kluwer Academic Publishers, Boston, Third Edition.

Butler, B. 1999. *The dynamics of cyberspace: Examining and modeling online social structure.* Unpublished Dissertation, Carnegie Mellon University, Pittsburgh, PA.

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

Butler, B.S. 2001. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research*, *12(4,)* 346-362.

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. 1996. Validating computational models. Working Paper, Pittsburgh, PA.

Cartwright, D., Zander, A. 1953. Group cohesiveness: Introduction. Cartwright, D., & Zander, A., eds. *Group Dynamics: Research and Theory*. Row Peterson, Evanston, IL.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Earlbaum Associates, Hillsdale, NJ.

Collins, N., & Miller, L. 1994. Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin, 116*(3), 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.

Csikszentmihalyi, M. 1997. *Finding flow: The psychology of engagement with everyday life*. Basic Books, New York.

Festinger, L. 1954. A theory of social comparison processes. *Human Relations, 7(2)*, 117-140.

Festinger, L. (1950). Informal social communication. *Psychological Review, 57*(5), 271-282.

Figallo, C. 1998. *Hosting web communities: Building relationships, increasing customer loyalty, and maintaining a competitive edge*. John Wiley & Sons, New York.

Fisher, D., Smith, M., & Welser, H. T. 2006. You are who you talk to: Detecting roles in Usenet newsgroups. In *Proceedings of the 39th Hawaii International Conference on System Sciences*, Waikoloa, Big Island, Hawaii, pp. 59b.

Foltz, P., & Dumais, S. (1992). Personalized information delivery: an analysis of information filtering methods. *Communications of the ACM*, *35*(12), 51-60.

Friedman, E., & Resnick, P. 2001. The social cost of cheap pseudonyms. Journal of Economics and *Management Strategy*, *10(2)* , 173-199.

Galbraith, J. 1973. *Designing complex organizations*. Addison-Wesley, Reading, MA.

Greenspan, R. 2004. Surfers prefer personalization. August 3, 2004. Retrieved from http://www.clickz.com/3389141 on January 7, 2009.

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(1)*, 68.

Harper, F., Li, X., Chen, Y., & Konstan, J. 1997. Social comparisons to motivate contributions to an online community. *Proceedings of Persuasive Technology*, Palo Alto, CA.

Harper, F.M., Frankowski, D., Drenner, S., Ren, Y., Kiesler, S., Terveen, L., Kraut, R.E., & Riedl, J.T. 2007. Talk amongst yourselves: Inviting users to participate in online conversations. *12th International Conference on Intelligent User Interfaces*, Honolulu, Hawaii, 62-71.

Herlocker, J.L., Konstan, J.A., Terveen, L.G., & Riedl, J.T. 2004. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, *22(1)*, 5-53.

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

Johnson, S.L., & Faraj, S. 2005. Preferential attachment and mutuality in electronic knowledge networks. *26th International Conference on Information Systems*, Las Vegas, NV. 287-299.

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(2)*, 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)*, 1-23.

Karau, S.J., & Williams, K.D. 1993. Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality & Social Psychology*, *65(4)*, 681-706.

Kollock, P. 1999. The economies of online cooperation: Gifts and public goods in cyberspace. In Smith, M.A., & Kollock, P., eds. *Communities in Cyberspace*. Routledge, London.

Konstan, J., Miller, B., Maltz, D., Herblocker, J., Gordon, L., & Riedl, J. (1997). GroupLens: applying collaborative filtering to Usenet news. *Communications of the ACM*, *40*(3), 77 - 87

Kraut, R. E., Wang, X., Butler, B., Joyce, E., & Burke, M. 2007. Building commitment and contribution in online groups through interaction. Working Paper. Carnegie Mellon University.

Lampe, C., & Johnston, E. 2005. Follow the (slash) dot: Effects of feedback on new members in an online community. *ACM Conference on Supporting Group Work*. ACM Press. 11-20.

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., Cosley, D., Frankowski, D., Terveen, L., Rashid, A.M., Resnick, P., & Kraut, R.E. 2005. Using social psychology to motivate contributions to online communities. *Journal of Computer Mediated Communication*, *10(4)*, article 10.

McKenna, K.Y.A., Green, A.S., & Gleason, M.E.J. 2002. Relationship formation on the Internet: What's the big attraction? *Journal of Social Issues*, *58(1)*, 9-31.

Ma, M., & Agarwal, R. 2007. Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Information Systems Research*, *18(1)*, 42.

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

Pew Internet. 2001. Online communities: Networks that nurture long-distance relationships and local ties. http://www.pewinternet.org/PPF/r/47/report_display.asp, accessed on May 1, 2007.

Postmes, T., & Spears, R. 2000. Refining the cognitive redefinition of the group: Deindividuation effects in common bond vs. common identity groups. In Postmes, T., Spears, R., Lea, M., Reicher, S., eds. *SIDE effects centre stage: Recent developments in studies of de-individuation in groups*. KNAW, Amsterdam, the Netherlands 63-78.

Preece, J. 2000. *Online communities: Designing usability, xupporting xociability*. Wiley, Chichester, England.

Preece, J., Nonnecke, B., & Andrews, D. 2004. The top 5 reasons for lurking: Improving community experiences for everyone. *Computers in Human Behavior*, *20(2)*, 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(5)*, 484-493.

Raghu, T. S., Jayaraman, B., & Rao, H.R. 2004. Toward an integration of agent-and activity-centric approaches in organizational process modeling: Incorporating incentive mechanisms. *Information Systems Research, 15(4)*, 316-335.

Rao, H. R., Chaudhury, A., & Chakka, M. 1995. Modeling team processes: Issues and a specific example. *Information Systems Research, 6(3)*, 255-285.

Ren, Y., Kraut, R.E., & Kiesler, S. 2007. Applying common identity and bond theory to design of online communities. *Organization Studies*, *28(3)*, 377-408.

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

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

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(1)*, 27-37.

Schafer, J., Konstan, J., & Riedl, J. 2001. E-commerce recommendation application. *Data mining and knowledge discovery*, Kluwer Academic Publishers.

Smith, M.A. 1999. Invisible crowds in cyberspace: Mapping the social structure of the Usenet. In Smith, M. A., & Kollock, P., eds. *Communities in Cyberspace*. Routledge, London.

Tam, K. Y., & Ho, S.Y. 2005. Web personalization as a persuasion strategy: An elaboration likelihood model perspective. *Information Systems Research*, *16(3)*, 271-291.

Vroom, V., Porter, L., & Lawler, E. 2005. Expectancy theories. In *Organizational behavior 1: Essential theories of motivation and leadership* (ed. Miner, J. B.), M.E. Sharpe, New York, 94-113.

Wasko, M., & Faraj, S. 2005. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, *29(1)*, 35-58.

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

Williams, K., Harkins, S. G., & Latane, B. 1981. Identifiability as a deterrent to social loafing: Two cheering experiments. *Journal of Personality & Social Psychology*, *40(2)*, 303-311.

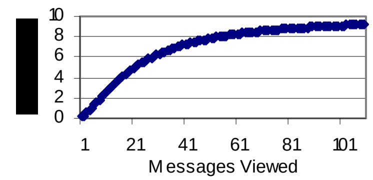
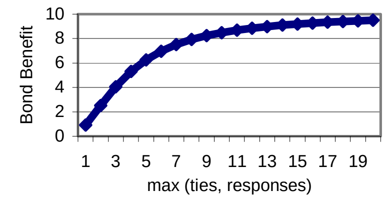
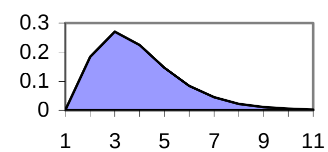
Worthen, B. Why most online communities fail? July 16, 2008. Retrieved on October 12, 2008 at http://blogs.wsj.com/biztech/2008/07/16/why-most-online-communities-fail/

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(2)*, 33-43.

Table 1. 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 latest messages followed by 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?	Conversations to join are jointly determined by preferential attachment, reciprocity, and match of personal interest

Table 2. Rationale, Rules, and Functions to Calculate Member Benefits

	Rationale	Rules	Function and Parameters
Information Benefit			
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 and Williams 1993)	InfoB_{share} is conditional on liking task or group, is greater when others under-contribute, and is greater when group size is smaller	$\begin{cases} 0, & if\ \max(IntrB_{rec}, SocB_{iden}, SocB_{bond}) < 3 \\ f(avg\ msg, group\ size), & otherwise \end{cases}$
Social Benefit			
From attachment to group ($\text{Soc}B_{iden}$)	Group identity (Hogg, 2000)	$SocB_{iden}$ is greater when agent's interests are similar to group interests	$f(\frac{count\,(viewed\ messages\ that\ match)}{count\,(viewed\ messages)})$
From attachment to members ($\text{Soc}B_{bond}$)	Interpersonal bonds (Sassenberg, 2002) Empirical studies of Usenet groups	$\text{Soc}B_{bond}$ is greater with repeated, mutual interactions, with immediate responses from other community members, and is a marginally decreasing function of the number of relationships an agent forms	
Other Benefit			
From recreation ($IntrB_{rec}$)	Intrinsic motivation Empirical studies of online behaviors (Cotte et al. 2006; Ridings and Gefen 2004)	IntrB_{rec} is a function of individual differences, distributed normally at the agent level, and distributed as a gamma distribution at the community level	

From reputation ($IntrB_{rep}$)	Incentive mechanisms (Wasko & Faraj, 2005)	When an individual member is among the topic 10% contributors	$f\left(\frac{self\ contribution}{\max contribution/10}\right)$

Table 3. 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 identity benefit, bonds benefit, or recreation 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-contribution*/
    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
```

Table 4. Summary of the Effects of Community-Level and Personalization Moderation

	Community-level moderation	**Personalized moderation**
Member commitment (logins)	• positive, significant • a greater effect in narrow-focus or high-traffic groups	• positive, significant • a greater effect in broad-focus or high-traffic groups
Community activity (posts)	• negative, not significant	• positive, significant • a greater effect in broad-focus or high-traffic groups
Benefit from information access	• positive, significant	• positive, significant • a greater effect in broad-focus or high-traffic groups
Benefit from interpersonal bonds	• negative, significant	• positive, significant • a greater effect in broad-focus or high-traffic groups

Figure 1. The Conceptual Framework for the Agent-Based Model

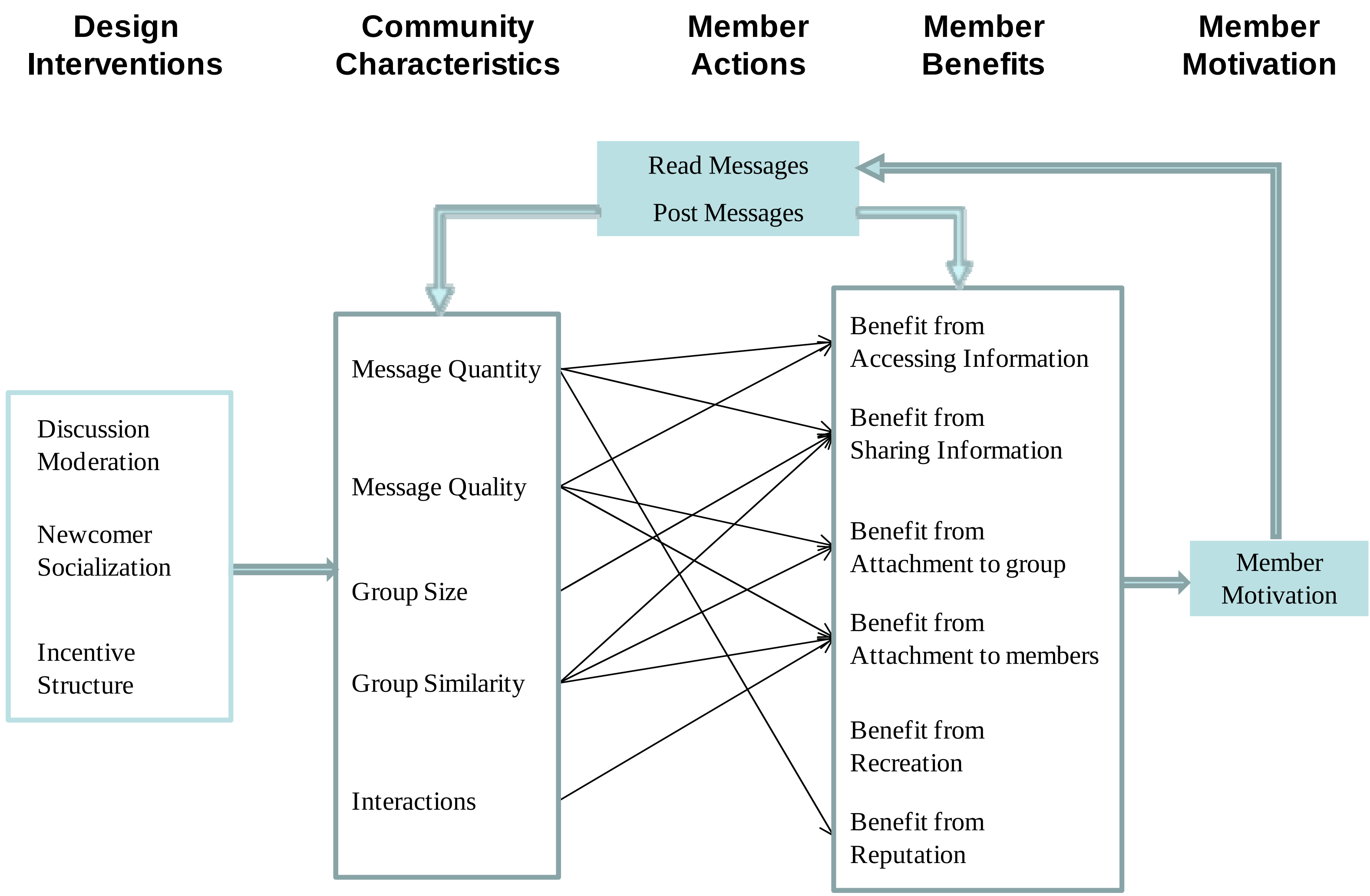


Figure 2. Comparison of Real (left) and Simulated (right) Data in Community Statistics

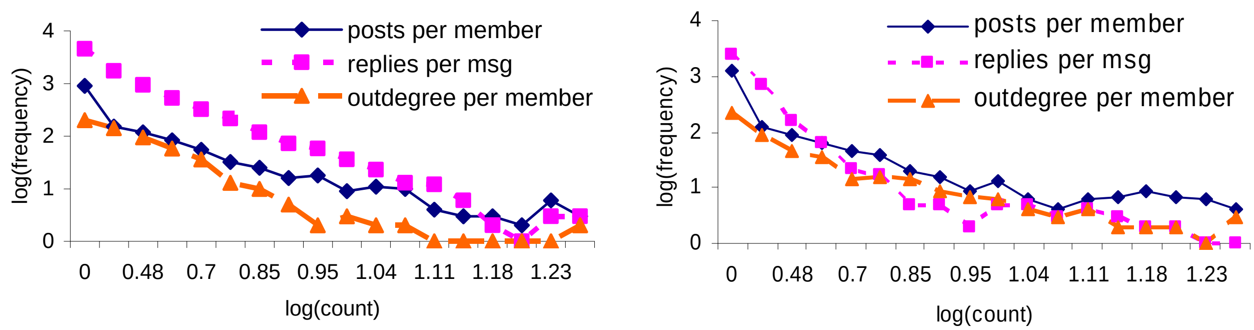


Figure 3. Comparison of Real (left) and Simulated (right) Data in Member Survival

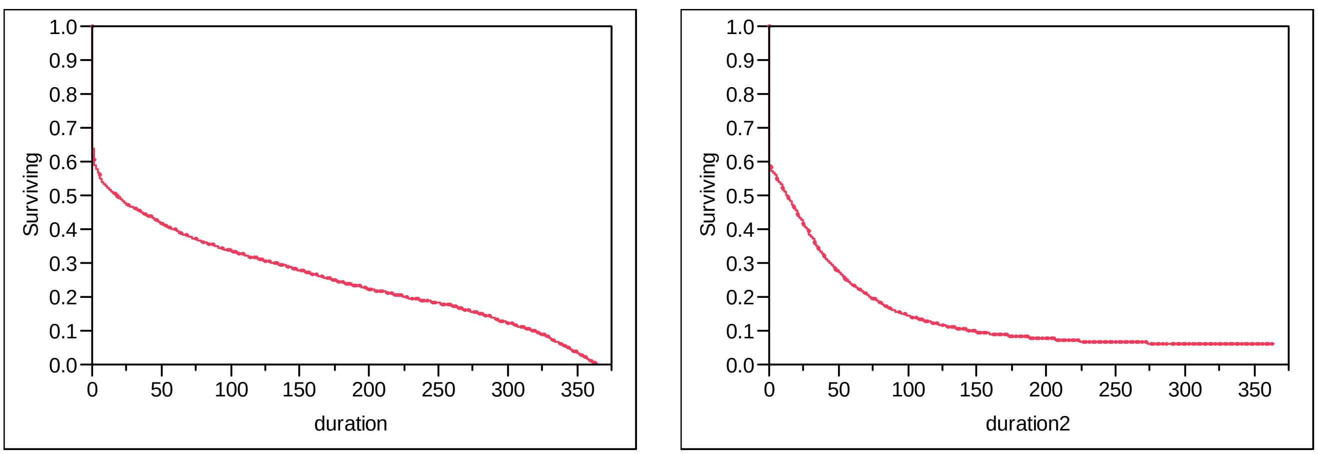


Figure 4. Effects of Moderation on Community Activity When Topical Breadth Varies

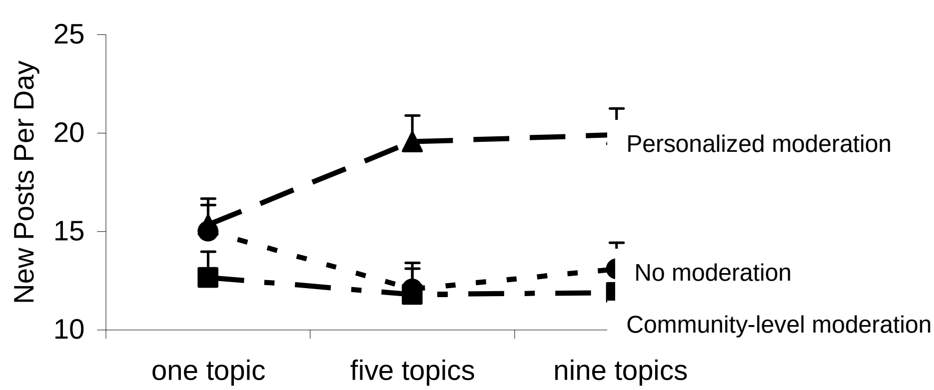


Figure 5. Effects of Moderation on Community Activity When Message Volume Varies

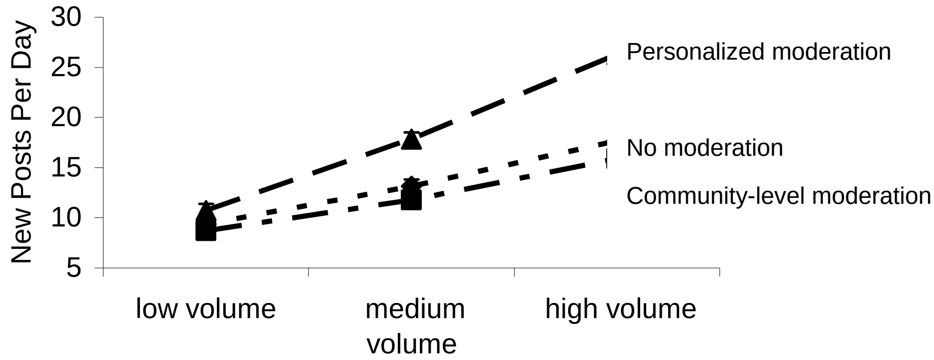


Figure 6. Effects of Moderation on Member Commitment When Topical Breadth Varies

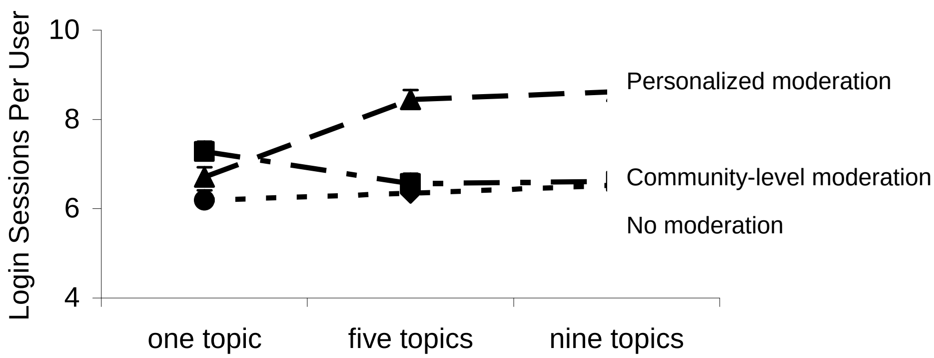


Figure 7. Effects of Moderation on Member Commitment When Message Volume Varies

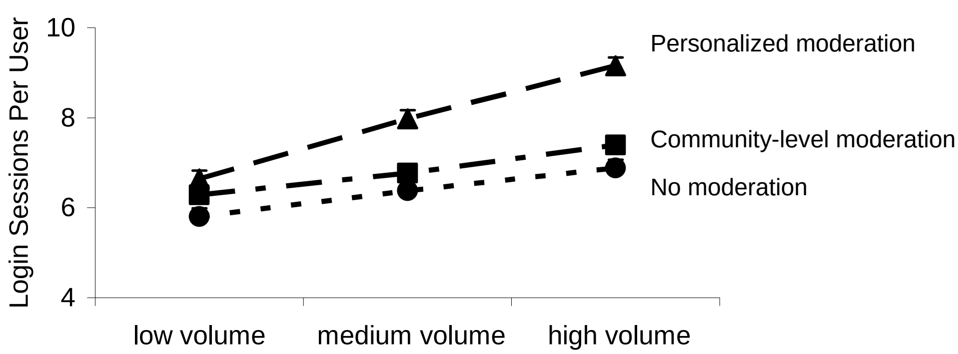


Figure 8. Effects of Moderation on Informational and Relational Benefit by Topic Breadth

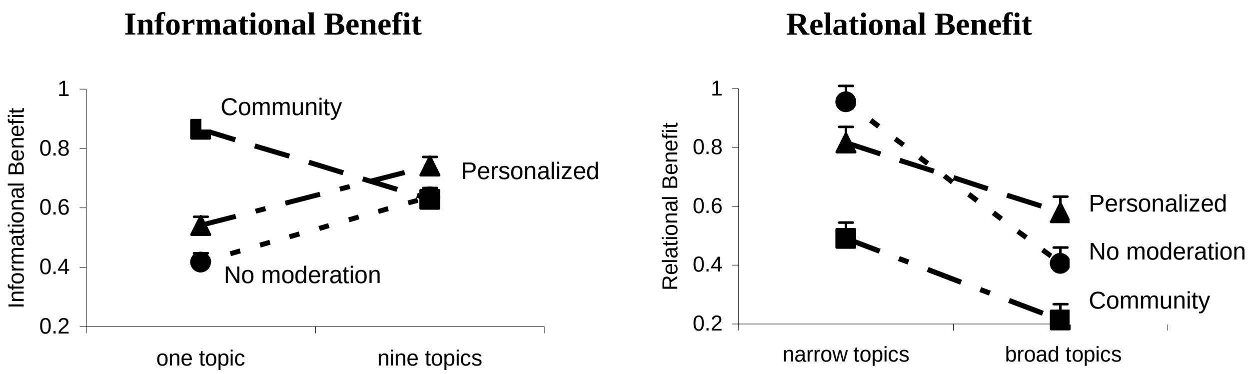
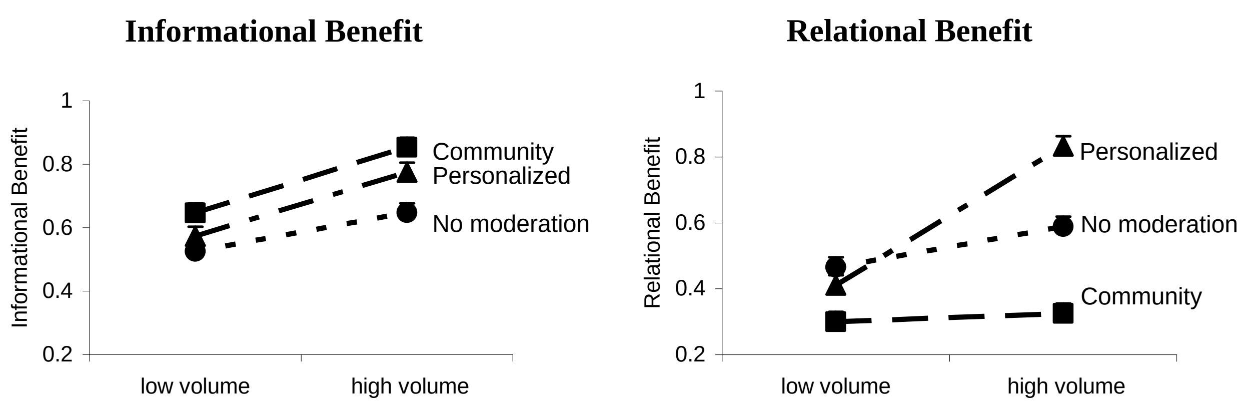


Figure 9. Effects of Moderation on Informational and Relational Benefit by Message Volume



Appendix: Sequences of Decisions an Agent Makes in a Simulated Day