

Membership Claims and Requests: Conversation-Level Newcomer Socialization Strategies in Online Groups

Small Group Research

41(1) 4–40

© The Author(s) 2010

Reprints and permission: <http://www.sagepub.com/journalsPermissions.nav>

DOI: 10.1177/1046496409351936

<http://sgr.sagepub.com>

 SAGE

Moira Burke,¹ Robert Kraut,¹
and Elisabeth Joyce²

Abstract

Early socialization experiences have a long-term impact on newcomers' satisfaction, performance, and intention to stay in a group. We know that newcomers proactively shape their own socialization, but we know little about the behavioral tactics they employ, or how the words they choose affect their acceptance by the group. The present article highlights three common conversational strategies of newcomers to online groups: (a) group-based membership claims, in which newcomers describe initial participation in the group; (b) identity-based membership claims, in which they describe their similarity to the group's focal social category; and (c) information requests, in which they ask for help. Using machine learning to identify these conversational strategies automatically in 12,000 newcomers' messages to approximately 100 online groups, we find that they are correlated with increased group responsiveness. We follow this analysis with two controlled field experiments to demonstrate that when individuals attest to previous group participation and make specific requests for information, community responsiveness increases, but claims of shared identity with the group have no impact.

¹Carnegie Mellon University, Pittsburgh, PA, USA

²Edinboro University of Pennsylvania, Edinboro, PA, USA

Corresponding Author:

Moira Burke, Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Email: moira@cmu.edu

Keywords

online communities, computer-mediated communication, newcomer socialization

Entrance to groups of all kinds—from corporate sales teams to fraternities to World of Warcraft guilds—requires a socialization process during which both the individual and the group determine their suitability for each other (Anderson, Riddle, & Martin, 1999; Levine, Moreland, & Ryan, 1998; Moreland & Levine, 1982; Reichers, 1987). Socialization affects long-term outcomes, including task performance, satisfaction, commitment, and the intention to remain in the group (Barge & Schlueter, 2004; Bauer, Bodner, Erdogan, Truxillo, & Tucker, 2007; Saks, Uggerslev, & Fassina, 2007). Recent literature has emphasized the role of the individual in the socialization process: Newcomers are not just passively indoctrinated; they proactively shape their own experiences (Miller & Jablin, 1991; Morrison, 1993, 2002; Wanberg & Kammeyer-Mueller, 2000). In particular, newcomers seek to build relationships with other members and the group as a whole, to gain feedback about their role in the group, and to perform cognitive self-management (Griffin, Colella, & Goparaju, 2002; Wanberg & Kammeyer-Mueller, 2000). These socialization moves occur in groups both online and off (Ahuja & Galvin, 2003).

Despite a large body of research on the proactive socialization tactics applied by newcomers to groups, we know very little about the concrete ways in which newcomers execute these socialization moves. Although we know, for example, that newcomers attempt to gauge a group's fit to their needs and to assess a group's potential acceptance of them (Moreland & Levine, 1982), we do not know the probes they use or the data they collect to make these assessments. For example, prospective employees often attempt to reduce uncertainty prior to their first interviews by contacting organization insiders, but we do not know the content of those contacts or how they affect integration into the group (Flanagin & Waldeck, 2004; Jablin, 2001).

Furthermore, much of this work has been influenced by Van Maanen and Schein's (1979) work and has relied on broad self-report instruments to measure socialization tactics (Jones, 1986; Mignerey, Rubin, & Gorden 1995). For example, Bartel, Wrzesniewski, and Wiesenfeld (2007) present survey data showing that marginal employees, such as newcomers or remote workers, make tentative membership claims by asking questions that demonstrate interest, making remarks that reveal insider knowledge, or simply declaring

their membership. Organizational commitment and identification are increased if group members ratify these overtures. Similarly, Mignerey et al. (1995) find that individuals' communication traits and attitudes are connected to their information- and feedback-seeking behaviors during socialization, but these traits do not directly explain socialization outcomes, perhaps because the surveys do not take into account the specific communication moves performed by the newcomers. The wider literature on organizational socialization tactics, reviewed by Bauer et al. (2007) and Saks et al. (2007), is similar, with most research relying on newcomers' self-reports of their tactics and revealing little about the behaviors through which they executed those strategies.

The present article fills this gap by examining three common conversational tactics used by newcomers to claim group membership and assessing whether the group responds positively. We examine *group-based membership claims*, in which newcomers describe their initial participation in the group, *identity-based membership claims*, in which they describe their similarity to the social category around which a particular group is organized, and *information requests*, in which they request help from the group, thereby implicitly claiming a right to membership benefits at the same time as receiving information about the group's willingness to accept them. We describe a methodological contribution, using machine learning to identify these tactics automatically in a sample of approximately 12,000 newcomers' messages, and follow the machine learning analysis with two controlled field experiments that demonstrate that group-based membership claims and specific information requests elicit membership grants, but identity-based membership claims do not.

Context of Online Groups

We use the context of online groups to delve into socialization tactics at a fine-grained level and to test principles of socialization first developed in traditional organizations. Clearly online groups differ from conventional groups on many dimensions: They have no shared physical space, they often have higher turnover, and their membership is less distinct. However, the considerable evidence that socialization practices across a broad set of social entities, including informal small groups, voluntary associations, employment organizations, and online groups, have comparable effects on important outcomes (including longevity, commitment, and satisfaction) suggests that online groups are an appropriate site to examine principles of socialization despite these differences (Bauer et al., 2007; Bryant, Forte, & Bruckman,

2005; Clary et al., 1998; Ducheneaut, 2006; Galegher, Sproull, & Kiesler, 1998; Saks et al., 2007; Von Krogh, Spaeth, Lakhani, & Hippel, 2003). Furthermore, online groups are interesting social phenomena in their own right for the benefits they provide members, including friendship, social support, entertainment, and information (Ridings & Gefen, 2004), and their exponential growth. One of the earliest networks of online groups, Usenet, had approximately 200,000 active groups, more than 9 million unique contributors, and 250 million messages in 2003 (Smith, 2003), whereas the newer social networking site Facebook has 250 million active members, half of whom log in daily (Facebook, 2009).

Challenges of Newcomers to Online Groups

Online, newcomers face special difficulties as a result of the diffuse, decentralized, and anonymous text-based interactions inherent in most online groups. As in conventional groups, they must evaluate the potential benefit of the group to decide whether membership is worth the effort of participation (Butler, 2001; Jablin, 2001). Without a face-to-face presence, shared organizational identity, or even much knowledge of a group's norms, newcomers must demonstrate a legitimate connection and commitment to the group (Galegher et al., 1998). In online discussion groups, they can estimate this future value by reading some of the archived conversations between existing members.

Although they can get some sense of the group by reading the archives, they can develop a more personalized estimate by posting a message to the group and assessing the replies, if any, that they receive. Previous research has shown that even a simple measure of community responsiveness—whether a poster gets a reply—is associated with increases in the poster's future participation (Baym, 1993; Joyce & Kraut, 2006), speed of contributions (Lampe & Johnston, 2005), and survival in the group (Arguello et al., 2006; Wang, Kraut, Butler, Joyce, & Burke, 2008). Replies to a newcomer signal the extent to which the group accepts them and leads to more committed behavior by the newcomer (Wang et al., 2008). Replies signal that the group believes the newcomer is a potentially valuable prospective member worth its attention, motivating the newcomer to reciprocate by writing more and replying to others. These conversations help transform newcomers into committed contributors. Posters interpret silence, on the other hand, as rudeness or unfriendliness (Cramton, 2002), which leads to withdrawal (Arguello et al., 2006).

Membership Claims and Grants

In many groups, membership boundaries are ambiguous, and new members are faced with uncertainty about their standing in the group (Mortensen & Hinds, 2002). Goffman (1959) proposes that individuals and groups seek to reduce uncertainty by defining a working consensus of membership status, built through an interdependent system of membership claims and grants in which peripheral members make the case that they belong and more established members validate these claims (Bartel & Dutton, 2001). These claims and grants consist of both words and actions, with both parties defining the relationship. The successful exchange of claims and grants leads newcomers to feel greater identification with and commitment to the group.

Membership claims take three forms: declarations, questions, and revelations. Declarations are verbal assertions by the newcomer that he or she is a member. They can be explicit statements of belonging or subtler cues, such as the use of the first-person plural (“we” or “us”) that implicitly define a newcomer’s connection to the group. Questions are a form of active inquiry crafted to convey an image of dedication and competence, an attempt to look more like an insider. Furthermore, in most groups members have privileged access to the group’s resources, and newcomers are making implicit claims of membership when they ask for these resources. Revelations are statements through which new participants reveal details about themselves, presenting their similarities with the group and demonstrating that they have an insider’s understanding of the group or shared values and ideology. Galegher et al. (1998) describe the legitimating function of these revelations in health support groups, through which newcomers disclose personal information relevant to the group’s topic and demonstrate their membership in the group-relevant category.

Claims alone do not make a member; those claims must be acknowledged and accepted by others in the form of grants. Grants may be verbal declarations of acceptance, such as welcoming messages or actions demonstrating that the newcomer is more inside than outside. For example, workplace managers may encourage a new employee to take on an important project as a way to communicate that the newcomer’s skills are valued. Involving the newcomer in highly visible or central activities provides evidence to both the newcomer and other members that the newcomer has been accepted (Myers & Oetzel, 2003). In online discussion groups, the most central activity is participation in an active conversation thread, and so the most direct way group members grant membership to newcomers is by replying to their messages.

Membership claiming and granting activities have been described at a general level (Bartel & Dutton, 2001; Bartel et al., 2007), but we know little about the specific forms these claims and grants take and the way their form influences their effectiveness. One of the contributions of this article is to extend the theory of membership claims and grants by specifying the types of conversational moves newcomers make to accomplish their claims. We present here an example of a newcomer to a breast cancer support group so that we can reveal in more detail the tactics she uses to elicit community response. The author's name has been changed, portions have been omitted for brevity, and italics have been added to emphasize the strategies. Otherwise, text remains as posted.

Hello everyone. My name is Angela. *I've been lurking* around your discussion group for a few weeks now. *Just reading and trying to soak in some knowledge* I guess. But~ I feel so dumb! *You all seem to know so much about this subject!* I'm so sorry that it has probably come from personal experience . . . *My mammogram* was on 12-6-02; . . . I was rather shocked that my gyno didn't take the time to at least speak with me and go over my radiology report. [Later] my radiologist said . . . there has been development of *a cluster of microcalcifications within the deep, slightly upper central left breast*. She stated that she felt there was at least a 70-75% chance that this was cancer. [But the original breast surgeon] thinks the chances of it being cancer are low—less than 30%. *Why the BIG difference in opinions???* So OK guys what do you think???

Is there anything you can tell me about these cluster Micro's, am I going about this alright or all wrong. Also, just out of curiosity, was that normal for my gyno to brush me off like that?

In this passage, Angela uses three types of membership claims—group-oriented claims, identity-oriented claims, and questions—to start a virtuous cycle of engagement in the group. In this case, members of the group sent five replies to her post, containing welcomes to the group, words of support, links to more information about her specific diagnosis, and recommendations to consult another physician. In the following sections, we describe these three types of claim in more detail.

Group-Oriented Membership Claims

Miller and Jablin (1991) describe information-seeking tactics used by newcomers, including observation of coworkers performing a task and eavesdropping

on coworkers' conversations. These tactics provide models that the newcomer can emulate, and ambient clues for learning norms and values. The equivalent of these strategies in online groups occurs when newcomers silently read the groups' messages before participating actively (Preece, Nonnecke, & Andrews, 2004).

Newcomers often refer to the time they have invested in learning about the group via group-oriented membership claims, which Rafaeli, Ravid, and Soroka (2004) refer to as *de-lurking* messages. In them, newcomers describe the effort they have invested to learn about the group (Galegher et al., 1998), signaling a desire to belong and demonstrating that they have been trying to learn more about the group and its norms before participating more actively (Nonnecke & Preece, 2000; Von Krogh et al., 2003). Baym (1993) found that de-lurking posts frequently announced the entrance of a new member into the community, and were often followed by *welcoming committee* posts, indicating that the community was accepting the self-proclaimed lurkers.

In the aforementioned example, Angela makes a group-oriented membership claim by stating that she has been "lurking" and "trying to soak up knowledge." She claims, therefore, to be a member of the group already, despite having been invisible. Generally, newcomers make these group-oriented membership claims by explicitly describing the period of investigation, by using terms such as "lurking" or "reading," a length of time, and a reference to the group. They can also be more subtle, as when Angela indicates her knowledge of the group by saying, "You all seem to know so much about this subject!"

By signaling their initial investment in the group through these claims, newcomers make themselves more desirable as group members (Branscombe, Spears, Ellemers, & Doosje, 2002; Noel, Wann, & Branscombe, 1995). Their words convey devotion to the community and concern for its welfare, shown in their respect for the existing work of other members and the intention to avoid wasting the group's limited attention by asking redundant questions. Existing members may even perceive lurking as a form of legitimate peripheral participation that enables novices to learn the ropes of the group (Lave & Wenger, 1990). These group-oriented introductions should increase the group's willingness to grant membership status to the newcomers, by responding to their questions and comments.

Hypothesis 1: Newcomers' use of group-oriented membership claims will increase the community's willingness to respond to them.

Identity-Oriented Membership Claims

Another socialization strategy that newcomers use is identity-oriented membership claims. Here, newcomers disclose personal information that indicates their connection to the larger social category from which the community draws its membership. For example, they might describe their treatment history in a diabetes group, their voting history in a political issue group, or their memories of Jerry Garcia records in a Grateful Dead fan group. In the passage above, Angela makes the identity-based membership claim that she is a breast cancer survivor, by describing her diagnosis and using group-specific language (e.g., “mammogram,” “my radiologist,” and “cluster Micro’s”). By doing so, she shows that she is not only similar to others in the group in having this diagnosis, but knows their vernacular. Both these strategies suggest that she is worthy of the group’s attention. By showing their membership in a common social category, newcomers should increase their likelihood of eliciting help from others in that category (Hogg & Terry, 2000).

Numerous studies demonstrate that self-disclosure both signals the strength of the relationship between individuals and is a mechanism to strengthen the relationship (Collins & Miller, 1994). Disclosure of personal information also signals newcomers’ desire to increase the closeness of their relationship to the group and others in it (Altman & Taylor, 1973; McKenna, Green, & Gleason, 2002). Emphasizing similarity to others in the group is a common technique to build a relationship (Byrne, 1997). Relationship building is another proactive newcomer socialization tactic associated with successful integration and retention (Griffin et al., 2002; Wanberg & Kammeyer-Mueller, 2000).

By claiming common category membership, signaling legitimacy, and attempting to build a social relationship with existing group members, newcomers using identity-based membership claims make themselves appear more worthy of community membership, and as a result, the community should be more willing to grant that membership by responding to them.

Hypothesis 2: Newcomers’ use of identity-oriented membership claims will increase the community’s willingness to respond to them.

Information Requests

To learn more about a group, newcomers also often use information-seeking tactics, which generally improve their fit and eventual commitment to the group (Bauer et al., 2007; Bauer & Green, 1998; Levine & Moreland, 1994;

Miller & Jablin, 1991). In the prior example, Angela makes several requests, both for medical information and for advice about her doctors. Questions, which are often the first attempt by newcomers in online groups to engage group members, serve two functions. First, if answered, they fulfill the newcomers' information needs. Second, whether or not the group answers these questions signals the community's willingness to form a relationship with the new participant. Indeed, many questions posted in online groups are not intended to elicit information as much as they are opening gambits intended to start a conversation. In many cases, these questions serve as probes that newcomers can use to assess the group, their fit with it, and whether the group will be welcoming to them. Thus the answers that Angela received to her questions allow her to predict the type of relationship and benefits she will get from group membership.

Even though face-to-face turn-taking norms imply questions should elicit answers (Sacks, Schegloff, & Jefferson, 1974), in online groups members have little obligation to respond to newcomers. Indeed, 40% of potential thread-starting messages in online discussion groups go unanswered (Joyce & Kraut, 2006; Smith, 2004). Therefore, information requests do more than trigger discourse rules. By making a request, a poster also implicitly claims membership to a group. Bartel and Dutton (2001) note that "becoming an active inquirer can convey an image of competence, interest, and dedication. Such qualities are more typical of organizational insiders than outsiders, and thus, support claims that one possesses membership status" (p. 122).

By using probes to assess their fit, signaling competence, and taking part in matters central to the group, newcomers who make requests implicitly claim membership status, making the community more likely to respond to them.

Hypothesis 3: Newcomers' use of requests will increase the likelihood that a community will respond to them.

We conducted three studies to investigate the relationship between these three socialization strategies and community responsiveness. The first is a correlational analysis of approximately 12,000 messages from 99 Usenet discussion groups, using machine learning techniques to detect these three socialization tactics automatically. This is followed by two experiments to establish causality by reposting previous messages to Usenet groups, experimentally adding or removing group-oriented membership claims, identity-oriented membership claims, and information requests.

Study 1: Correlational Analysis of 12,000 Usenet Messages

The goals of Study 1 were to determine the relationship between newcomers' socialization tactics and community responsiveness and to demonstrate the value of machine-learning techniques to identify these tactics automatically in a large corpus.

Data Collection

The sample was drawn from all Usenet groups having at least four posts per week from June 2003 to February 2005. We randomly selected 25 groups that met the activity criteria from each of the following categories: health support, technical, hobby, and political issues. Health support topics include asthma, epilepsy, breast cancer, and food allergies. Technical groups include C programming, civil engineering, and Windows NT security. Hobby groups include quilting, the Grateful Dead, and vegetarian cooking. Issue groups include gun rights, economics, and agnosticism. The sample originally included 100 groups, but one group was excluded because we could not access its text archive.

Approximately 2 million messages were posted to the 99 groups during the focal period. We randomly sampled up to 500 first-in-thread messages by unique authors from each group. We focused on messages that were first in their threads because they were potential conversation starters rather than replies to ongoing conversations. Some groups had fewer than 500 messages meeting these criteria, and cross-posted duplicates and those written by non-newcomers were removed, resulting in a sample of 11,889 messages. A total of 61% of the messages received a reply. Microsoft's Netscan project (Smith, 2004) provided metadata, including the total number of messages posted to a group on a given day, dates of an individual's first and last posts, and the number of replies a message received.

Measures

Group-oriented membership claims. Two human judges hand-coded a set of 360 messages from discussion groups not in the sample, giving a binary score indicating the presence or absence of a group-oriented claim. Interrater reliability was good (Cohen's $\kappa = .93$), and a third judge settled any disagreements. Additionally, two dictionaries were created based on the judges' descriptions of the most informative words for identifying introductions: One

was a dictionary of *lurking* terms, including lurk, read, follow, hear, observe, join, learn, post, listen, or browse; the second was for *group* terms, including: you, you guys, y'all, u, newsgroup, ng, group, list, site, discuss, talk, board, post, thread, message, online, here. The human codes were used as the gold standard for a machine-learning model in the TagHelperTools application (Dönmez, Rosé, Stegmann, Weinberger, & Fischer, 2005). TagHelperTools treats each message as a set of features, including all individual words in the message (e.g., "hello," "please"), all bigrams (pairs of words, e.g., "thank you"), the presence of words from the lurking and group dictionaries, and all part-of-speech bigrams (such as a past participle followed by a present participle). This approach identifies patterns in tense and phrasing, even when writers have different vocabularies. From these features, TagHelper generated a binary decision tree that most accurately differentiated between the 121 messages in the training data containing group-oriented membership claims and the 239 messages that did not (see Figure 1). Features that appeared fewer than two times were excluded to prevent the model from overfitting to rare words. Header text, quoted text from previous messages, and signature blocks were removed before processing using the Jangada application (Carvalho & Cohen, 2004). All words were stemmed, so that variants of the same word were considered a single feature (e.g., "read," "reader," and "reading"), and noncontent functional words (e.g., "this" or "and") were removed after part-of-speech tagging was complete. This *bag of words* approach does not distinguish between phrases such as "I've been lurking" and "I hate people who lurk;" it simply notes that both contain a term from the lurking dictionary. However, in spite of the noisy results, human review of the classification results indicates the model generally performs well. These probabilistic word-based approaches have successfully been used to explore many social psychological phenomena, including recovery after trauma (Cohn, Mehl, & Pennebaker, 2004), gender differences (Newman, Groom, Handelman, & Pennebaker, 2008), and lying (Newman, Pennebaker, Berry, & Richards, 2003).

Figure 1 shows the resulting tree for group-based membership claims. If the message has at least one term from the lurking and group dictionaries, the classifier checks for use of the perfect progressive tense (the past of "to be" followed by a present participle, e.g., "been reading"). If this part-of-speech pair is present, the message is classified as having a group-oriented introduction. The part-of-speech step is consistent with common sense: It indicates an action transpiring over time in the past. Instances of the present perfect progressive ("have been reading," for example) suggest that the action started in the past and continues through the present, and that the individual began

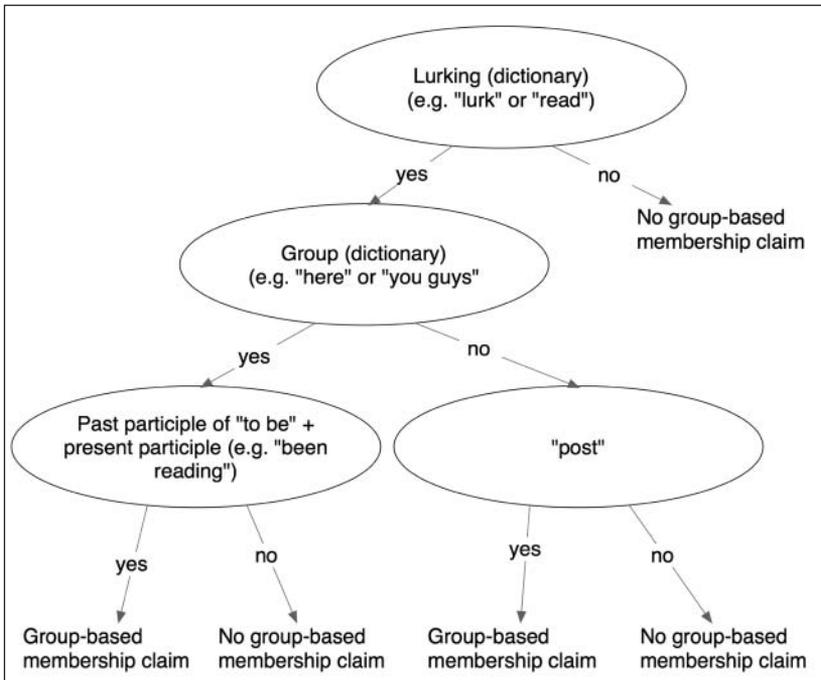


Figure 1. Model of group-oriented membership claims

reading messages posted in the group at some past date and still does so. If the message had a lurking term but no group term, the classifier looks for any form of the word “post.” If this term is present, the message is classified as having a group-oriented membership claim. The model was 92% accurate with a Cohen’s κ of .75 under 10-fold cross-validation: The model trained on 90% of the data and tested on the remaining 10%, 10 different times (Altman, 1991). The resulting model was applied to the 11,889 messages in the main sample, giving a dummy variable, *has group-oriented membership claim*, a score of 0 or 1.

Identity-oriented membership claims. Whereas group-oriented membership claims tend to be similar across groups, identity-oriented membership claims rely on group-specific vocabulary and therefore are more difficult to detect automatically. Human judges looking for identity-oriented membership claims found that they often include personal statements of self-identification, particularly self-labeling (e.g., “I have been a cow breeder, so I understand the principles

of selection and improving the breed”). They often include the length of time the author has held the identity (e.g., “I am an old and seasoned NCO with nearly 20 years under my belt”). Newcomers often start their conversations with a group by describing attempts to solve a relevant problem, such as searching for special quilting tools at sewing shops or trying multiple studios to find a good yoga instructor. When these problem descriptions provide relevant details to allow other readers to troubleshoot their problem but do not refer to their relationship to the domain, we do not consider them identity-based membership claims (e.g., “I’m running apache 2 on redhat 9”). They are identity-based membership claims, though, when the authors describe their personal relationship to the domain (e.g., “I’ve been running Linux boxes in my home office for two years.”).

To classify a message as containing an identity-oriented membership claim, we used a two-step process that first looked for general first-person terms describing experiences and then calculated how on-topic the message’s vocabulary was compared with that of other messages in the group (see Figure 2). Using this procedure means that a disclosure of medical history in a C++ programming group, although personal, is categorized as off-topic for the group. This type of instance would not be classified as an identity-oriented introduction as it does not serve to legitimate the newcomer. Because human judges determined that identity-oriented membership claims typically occur near the beginning of messages, we applied these rules only to the first 500 characters of each message. A training set of 714 human-coded sentences from other similar discussion groups was created. Approximately 20% (145) of those sentences had an identity-oriented membership claim. Two dictionaries were again defined: One for *time* words that would indicate length of connection to the topic, including variants of: month, week, year, day, awhile, since I, forever, decade, long time, current, now, recent, once, when, always, ever, since, time, ago; the second was for self-identification words and bigrams: I’m a, am a, I am, I try, have been, I have, I live, I was, I started, myself, a fan, a believer, my.

First, the model looks for references to time, terms from the self-identification dictionary, and/or the category of newsgroup. One word in the time dictionary was surprising: The model classified messages with the word “day” in them as not having identity-based claims. Human inspection of the classification revealed that these messages were often spam, such as “I made \$4000 in 30 days from home!” Additional spam messages were identified with a spam application in the overall membership-granting model (see control variables). The model also singled out technical newsgroups, only requiring a reference to time, possibly because these references are less

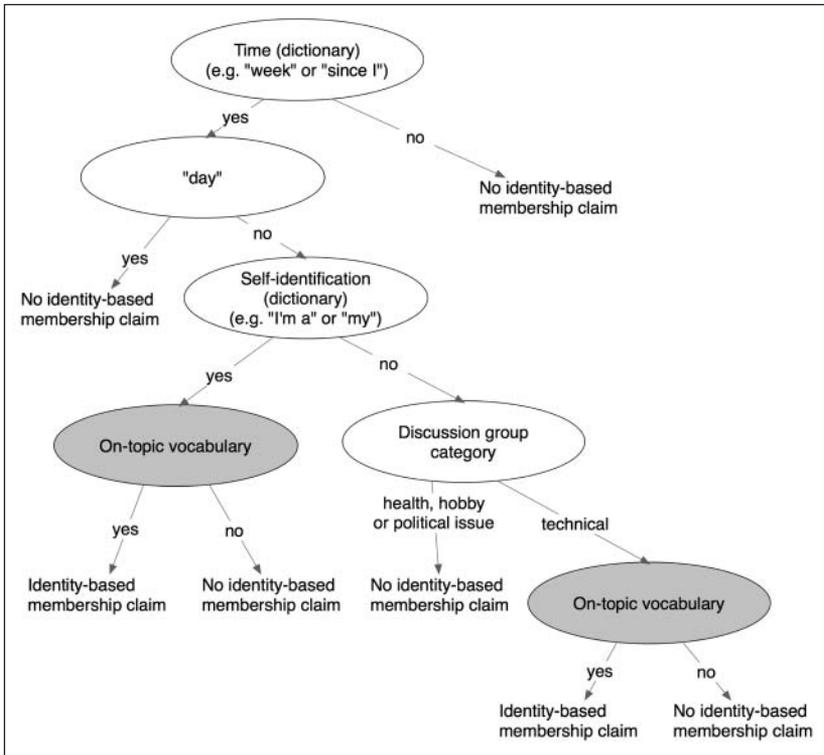


Figure 2. Model of identity-oriented membership claims

Note: Shaded ovals represent the second stage of the analysis, in which vocabulary topicality is measured.

common in nontechnical Q&A (question and answer) messages. This model correctly classified 85% of the training messages ($\kappa = .44$). This κ shows moderate agreement with the human coders (Landis & Koch, 1977). Although far from perfect, it is much better than chance agreement (where $\kappa = 0$), and leads to conservative hypothesis testing.

The second step in modeling identity-oriented membership claims was to determine how similar the vocabulary in the message was to the vocabulary of the group to which it was posted. Messages identified by the decision tree as having an identity-oriented membership claim were passed in their entirety through a topicality checker, which generated an index of average document frequency (ADF). ADF measures the extent to which the language in the focal message is widely used in the group. It is defined as the number of

messages in the discussion group in which each content word of the focal message appears one or more times, divided by the number of words in the focal message and total messages in the group. Before passing the messages to the ADF process, words were stemmed and functional terms (i.e., the, this, is) were removed. Topicality scores ranged from 0 to 70, with a mean of 6.7 and a standard deviation ($SD = 7.4$). We set a dummy variable, *has identity-oriented membership claim*, to 1 if the decision tree indicated identity-oriented membership terms and the ADF for the message was greater than the group's mean. Examples of identity-based membership claims identified by this procedure include the following: "I've lost complete custody of my son a year after my divorce 3 years ago" (alt.support.parents.with-custody), and "I've been a lover of jazz for a long time and just recently started playing around with improvisation on the piano" (rec.music.bluenote.blues).

Information requests. A request for information may include question marks, interrogatives, reversed subject and verb, indirect requests (e.g., "I want," "I'm looking for"), or references to help (e.g., "suggestions," "advice," "recommendations"). For information requests, a human-coded training set of 1,011 messages from two discussion groups was used as the training data. The Boosted Stump algorithm in the Minorthird machine learning package (Cohen, 2004) resulted in a κ of .61, performing nearly as well as a human judge and far better than chance. A dummy variable, *has request*, is 1 if Minorthird classified the message as containing a request, and 0 otherwise.

The algorithm does not provide an easily readable decision tree, but produces a list of the most informative features that predict whether a message is a positive or negative example of an information request. Some of the best positive indicators of requests included the words: *thanks, anyone, I, wondering, what, help, seem, ?*. Each of these words makes sense in the context of requests. For example, "thanks" often occurs in requests as in phrases such as, "thanks in advance," which show gratitude for the expected response. "Anyone" is a way to address the entire group, used as in phrases such as, "Does anyone know . . .," when a newcomer does not know which specific person to ask. Additional interrogatives (e.g., how, where, who) were most likely not found to be informative predictors of requests because their high frequency reduced their information value. Negative indicators of requests, as determined by the machine learning algorithm, included terms that may be more correlated with statements of opinion, rather than requests: *we, you, their, f**k, see, !*.

Control variables. Factors previously determined to affect the likelihood of getting a reply were included as controls in this analysis. Control variables, listed in Table 1, include the group type, message traffic the day the focal

Table 1. Number of Replies as a Function of Membership Claims and Requests

	Incidence Report Ratio	Standard Error	Expected No. of Replies
Intercept ^a			1.49
Has group-based membership claim	1.38***	0.06	2.05
Group-based claim × political issue group	0.89	0.07	1.32
Group-based claim × health support group	1.45***	0.09	2.16
Group-based claim × technical group	0.82**	0.05	1.22
Has identity-based membership claim	1.36***	0.04	2.03
Identity-based claim × political issue group	0.96	0.05	1.43
Identity-based claim × health support group	1.33***	0.06	1.98
Identity-based claim × technical group	0.91*	0.03	1.36
Has request	1.40***	0.03	2.09
Controls			
Political issue group ^b	0.65***	0.03	0.96
Health support group ^b	1.12***	0.04	1.67
Technical group ^b	1.22***	0.04	1.82
Is likely spam	0.37***	0.08	0.55
Is cross-posted	1.84***	0.06	2.74
Message length; Ln(words)	0.99	0.01	—
Group traffic on focal day; Ln(messages)	0.98	0.01	—
Average daily group traffic that year; Ln(messages)	1.23***	0.01	—

Note: All variables are binary unless otherwise indicated. Continuous variables were logged and centered. $N = 11,889$ messages.

a. The intercept represents the grand mean across all group types, with no membership claims or requests and all continuous variables at their mean values.

b. Discussion group types were converted to binary effect coding variables, with hobby groups as the omitted category. The coefficients represent the difference between that type's average and the grand mean across all four group types. Thus, messages posted to political issue groups received 65% of the mean number of replies. Support groups, on the other hand, received 12% more replies than average.

* $p < .05$. ** $p < .01$. *** $p < .001$.

message was sent and over the year, message length, and whether the message was cross-posted to multiple groups (Arguello et al., 2006; Fisher, Smith, & Welser, 2006; Jones, Rafaeli, & Ravid, 2004; Wasko & Faraj, 2005). Spam messages are unlikely to include either group-oriented or identity-oriented membership claims or requests and are also unlikely to receive a response, so we used SpamAssassin, a commercial spam filter, to generate a dummy variable, *spam*, which equals 1 if the message is likely spam, and 0 otherwise. Continuous variables representing word count, focal day traffic,

and annual traffic were logged and then standardized to have mean values of 0 and *SDs* of 1.

Results and Discussion

To examine the effects of group- and identity-oriented membership claims and information requests on the group's membership grants to newcomers, we use negative binomial regression to predict number of replies each message received. Table 1 shows the results.

Overall, messages received a mean of 2.0 responses ($SD = 3.6$), and the conversational content had a strong impact on how much the community responded. Table 1 presents the number of replies a message would be expected to receive when setting each variable to 1 (for binary variables) or increasing it by one standard deviation (for continuous variables), holding all other independent variables constant. Incidence report ratios, or the factor by which reply counts change, are also included. We also include the interactions of introduction type or requests with group type in Table 1; the exploratory analyses indicated that their inclusion improved the model.

Posts classified as containing group-oriented membership claims received 38% more replies than those without ($p < .001$). Furthermore, the effect differed depending on the kind of discussion group. Group-oriented membership claims increased reply rates in support groups by 45% ($p < .001$) but actually reduced reply rates in technical groups by 18% ($p < .01$) compared with the average. Identity-based membership claims also increased reply rates by 36% ($p < .001$), particularly in support groups (33%, $p < .001$), but reduced community responsiveness in technical issue groups by 9% ($p < .05$). This is consistent with normative behavior to establish legitimacy that dictates self-introductions in health support groups (Galegher et al., 1998) and concise expression in technical groups (Raymond & Moen, 2006). Requests also increased the number of replies across all types of groups by 40% ($p < .001$).

Although the present work examines only the first move in what may become a virtuous cycle, other research has shown that replies increase newcomers' commitment to the group substantially (Wang et al., 2008). A post hoc analysis of the current data shows that newcomers who received even a single reply to their first message were far more likely to post again and remain in the group longer than those who did not receive a reply. Figure 3 graphs a survival analysis, showing the percentage of newcomers remaining in the group over time within the 20-month observation period. Only 16.7% of newcomers who failed to receive a reply ever posted again in their group,

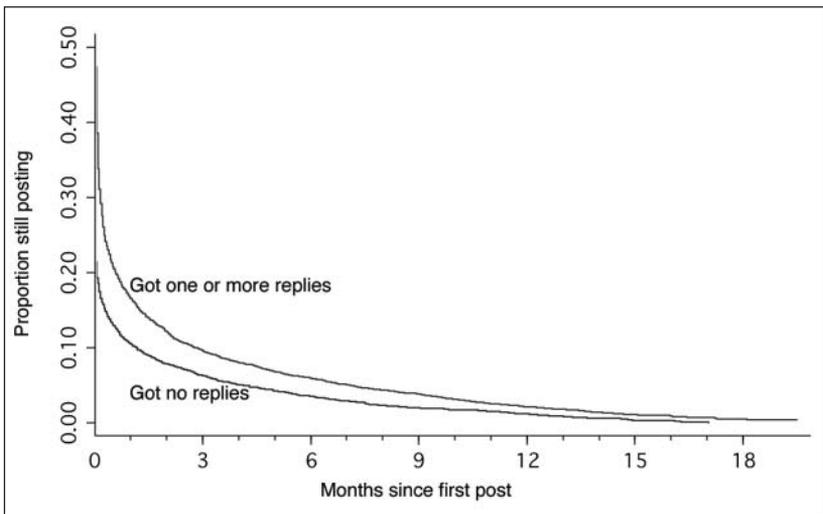


Figure 3. Survival analysis of newcomers after their first post, split by those who received at least one reply and those who did not

Note: Time is aligned across users so that time zero represents the day of the newcomer's first message.

whereas 45.5% of those who received a reply came back ($p < 0.001$). The difference persists for at least a year.

To summarize, Study 1 shows that the socialization strategies of membership claims and requests have a significant impact on community responsiveness in online groups. After controlling for numerous low-level characteristics of a message and the environment in which it was posted—including newsgroup traffic, message length, and spam—the conversational strategies of the message author appear to affect the number of responses. These results are more impressive because the conversational strategies were measured automatically, using procedures that were only modestly accurate, with the chance-adjusted agreement between human judges (κ) ranging from .44 to .75.

Although it is plausible that the relationship between the explanatory variables and the outcome is causal, and thus can be the basis of interventions to improve success in online groups, the data are correlational. It may be that other characteristics of the people who post messages with membership claims or requests or other features of these types of messages are responsible for the higher response rates. Therefore, Studies 2 and 3 were designed to establish the causal influence of membership claims and requests.

Table 2. Base Message With Group- and Identity-Based Membership Manipulations

Membership Claim Type	Message Text
None	Original from alt.support.cerebral-palsy Subject: Neuromove Anyone had any experience with this device? http://www.neuromove.com/
Group-based	<i>I've been reading here for the last month and am ready to jump in.</i> Anyone had any experience with this device? http://www.neuromove.com/
Identity-based	<i>My son has cerebral-palsy and I've been looking for options.</i> Anyone had any experience with this device? http://www.neuromove.com/

Study 2: Introduction and Disclosure Manipulations

In Study 2, previously posted messages from 93 Usenet groups were experimentally manipulated to add or remove group- and identity-based membership claims and then reposted to the original groups to measure the impact of the manipulations on the community's response. The high turnover rate—only 28% of posters in 1 month remain the following month (Wang, 2007)—suggests that there is little likelihood that the author or readers of the original message would observe the reposted message.

Method

Of the 99 groups from Study 1, 93 were still active in July 2006. From each of the 93 groups, two or three potential thread-starting messages (i.e., not replies to existing threads) at least 1 year old were randomly selected as base messages from the Google Groups Web site. Spam, FAQ reminders, outdated topics, and obvious flaming were filtered out. For each of the base messages, three versions were derived, including the original message and one was randomly selected to repost to the newsgroup: (a) a version with no membership claim, (b) one with a group-oriented membership claim, and (c) one with an identity-oriented membership claim. For simplicity, we did not manipulate both kinds of claims in the same message. One of the three versions was the original message. Table 2 shows three versions of a base message. The experimental messages were reposted using new accounts, and thus from the group's point of view were posted by newcomers.

To check the validity of the manipulations, two independent judges coded the messages for the presence of the two kinds of membership claims using the same guidelines as those used to train the machine learning algorithm in Study 1. Cohen's κ for interrater reliability was .87 for group-based claims and .91 for identity-based claims. Judges also ensured that manipulations in which claims were removed still maintained the nature of the original messages.

The experiment employed a 4 (group category: support, technical, issue, hobby) \times 3 (membership claim: none, group, identity) design across 93 groups, with 18 to 25 messages for each of the 12 combinations.¹ The dependent variable was the number of replies received within 1 week. A preliminary analysis of more than 200 million threads from the Netscan database shows that 96% of thread-starting messages that eventually receive a reply do so within 24 hours, so a week-long window is adequate to count replies. The messages were posted in counterbalanced order, approximately 15 messages each day for 3 weeks.

Research involving interactions with online groups requires special ethical considerations, and so the experiment included several precautions to protect participants. To ensure that privacy norms were not violated, the experiment included only large public groups with no registration and high turnover ($M = 72\%$ per month) so that the appearance of newcomers repeating ideas is common. Replies in which the reposting was noticed were generally neutral in tone and showed that the reposting did not cause trouble for the group (see below). To ensure that the activity would not noticeably affect the character of a group, the messages comprised a small percentage of the group's traffic for the week it was posted (median = 4.1% across groups). Original author names were changed and personal URLs were removed from signature blocks. The University's Institutional Review Board approved the research as involving minimal risk.

Results and Discussion

The measure of community responsiveness in Study 2 was the number of replies. Of the 263 messages posted, only 6 posts (0.023%) were recognized as reposts of previous messages. Those messages received a total of 69 replies, 17 of which related to reposting. Those 17 replies were generally neutral (e.g., "Someone asked exactly this question a year ago and no one had an answer then") and were excluded from analysis to avoid artificially inflated reply counts. Four of the five groups that noticed the reposting were hobby groups (including one group that caught both reposted messages),

Table 3. Number of Replies as a Function of Membership Claims

	Incidence Report Ratio	Standard Error	Expected No. of Replies
Intercept ^a			1.84
Has group-based membership claim	1.46*	0.28	2.69
Has identity-based membership claim	1.03	0.21	1.90
Controls			
Political issue group ^b	0.52***	0.08	0.96
Health support group ^b	1.27	0.17	2.34
Technical group ^b	1.22	0.16	2.24
Group traffic that month; Ln(messages)	1.42***	0.06	—
Original reply count; Ln(messages)	1.19***	0.05	—

Note: All variables are binary unless otherwise indicated. Continuous variables were logged and centered. $N = 257$ messages.

a. The intercept represents the grand mean across all group types, with no membership claims and all continuous variables at their mean values.

b. Discussion group categories were converted to binary effect coding variables, with hobby groups omitted.

* $p < .05$. ** $p < .01$. *** $p < .001$.

suggesting that hobby groups might have a lower turnover rate or more robust collective memory than other group types.

Table 3 shows a model of the number of replies as a function of the socialization strategies: manipulated group- and identity-based membership claims. In addition to the independent variables, *has group-based membership claim* and *has identity-based membership claim*, the model controls for the number of replies the message originally received, which can be treated as a proxy for how interesting or compelling the message topic and style were, the traffic to that newsgroup that month (number of messages, logged, and standardized), and the kind of group to which it was posted. We used effects coding for group type, so the coefficients for group type (hobby, issue, support, and technical, with hobby as the omitted type) represent the difference between that type's average and the grand mean across all four types. The base condition, representing the grand mean across all group types, with no membership claims and all continuous variables at their mean values received 1.84 responses. Adding a group-based membership claim increased the number of replies by 46% ($p < .05$). Adding an identity-based claim did not significantly affect the number of replies. A likelihood ratio test indicates that adding interactions between membership claims and group type did not improve the model, $\chi^2(6, N = 257) = 10.45, p > .10$, and so interactions are

not included. As expected, the number of replies the message originally received—the degree to which it contained material to provoke a group's interest—was significantly correlated with the number of replies after reposting, and groups with greater traffic also replied more. As in Study 1, issue groups received 48% fewer replies on average than the other groups ($p < .001$), and support groups had marginally more replies ($p = .08$), but because of the small sample size, no difference was found in responsiveness between technical groups and the others. As a validity check, adding the day of week of the repost and whether the message was original text or a manipulation did not change the model results.

In summary, Study 2 showed that group-based membership claims increased reply counts by 46%, independent of message content and context. Unlike in Study 1, identity-based membership claims did not have a significant impact. This failure to replicate these results has several interpretations, including that the phrasing of the membership claim does not cause increased granting by the group, but rather other underlying variables may come into play. Other possible interpretations for this lack of results are described in the general discussion.

Study 3: Request Manipulations

Study 3 tests the causal relationship between requests and community responsiveness, and teases out difference between open-ended and specific requests. In their model of newcomer information-seeking tactics, Miller and Jablin (1991) describe different levels of directness, with indirect requests—hints—common in situations of high uncertainty or high social cost. Compared with replying to an existing thread, starting a conversation in an online discussion group is more uncertain and imposes a greater burden on the members. Posters attempt to mitigate this burden by using linguistic politeness strategies of indirectness, either by not making a request at all or by using open-ended language allowing for multiple interpretations of the request (Brown & Levinson, 1987). The post, “Anybody else taken this medication?” is an open-ended request in that it does not specify the asker's desire, whether for sympathy, details of side effects, or personal stories from others with a similar diagnosis. On the other hand, the post, “What side effects can I expect from this medication?” is far more direct, indicating exactly what the author seeks. This difference in request directness may affect the group's response. According to politeness theory, the first request is more polite than the second because it leaves room for ambiguity.

Although newcomers may use indirect requests for greater politeness and to relieve social stress, they may paradoxically increase the burden on the group and reduce its willingness to respond to their open-ended and vague posts (Francik & Clark, 1985; Linde, 1998). The clarity of a direct, specific request might require less from the group, even if it is less polite. Direct requests and questions are calls to action; they make clear what the newcomer hopes to get from the group. Netiquette pages, such as Raymond and Moen's (2006) *How to Ask Questions the Smart Way*, have detailed guidelines for making requests effectively:

Open-ended questions tend to be perceived as open-ended time sinks. Those people most likely to be able to give you a useful answer are also the busiest people . . . People like that tend to be allergic to open-ended time sinks. (<http://www.catb.org/~esr/faqs/smart-questions.html>)

Linde (1998) also found that calls for action in face-to-face conversation are less successful if they contain vague or mitigating language such as "would," "could," or "please." Therefore, to determine the causal impact of requests, and the relative impact of direct versus indirect requests, we performed a third experiment, using the same protocol as Study 2. Given the literature on the effectiveness of specific requests and existing netiquette pages encouraging posters to be specific, we predict that specific requests will increase community responsiveness.

Hypothesis 3a: Newcomers who make specific requests for information will elicit greater community responsiveness than will newcomers who use vague, open-ended requests or make no requests.

Method

As in Study 2, three messages from each newsgroup were randomly selected such that they were the first message in their threads and not obviously spam, an FAQ reminder, outdated topic, or flame. In Study 3, new newsgroups were selected for a total of 100, stratified across the four group types, to replace groups that had dropped below the minimum traffic threshold of 50 messages per week. Each base message was randomly assigned to one of three request conditions: no request, a specific request, or an open-ended request. One judge created the manipulated version of the message (or left it alone when the original version matched its randomly assigned condition) and a second judge blindly coded the message as a manipulation check. Where judges

disagreed, the manipulation was strengthened until both of them agreed the message fit the request condition.

Specific requests were phrases in which it is clear what the requestor needs (e.g., “Will it fit in a 172?,” “Where can I find this quilting block?”). *Open-ended requests*, however, demonstrate that the author is requesting something from the group, but without making the desired information explicit (e.g., “Has anyone heard of this [quilting] block?” in which we do not know if the author is trying to find it or wants to know if it is pretty, or “I’m wondering if anyone has experienced this?” in which we do not know if the author is looking for sympathy or information). Open-ended requests that are clarified by other text in the message body are counted as specific, as is a message with both a specific and open-ended request. *No request* messages do not show any requests of the group, though the author may still be hoping to elicit conversation, for example, “I’ve been reading here frequently, and thought some of you might enjoy this. If you have to suffer through four more years, might as well have a smile on your face. Satiric news, updated every week [Link URL omitted].” Phrases such as “Ibuprofen?” or “George Bush?” are not requests; nor are rhetorical questions. Table 4 shows three versions of a base message.

The experiment employed a 4 (group category: support, technical, issue, hobby) \times 3 (request: none, open-ended, specific) design across 100 groups. The dependent variable was the number of replies received within 1 week. The 300 messages were posted in counterbalanced order, approximately 20 messages each day for 2 weeks. Three messages were observed to be reposts and were removed from the analysis, and as in Study 2, community response in those cases was neutral.

Results

Table 5 shows the results of a negative binomial regression on the dependent variable, number of replies, controlling for the same contextual variables as in Study 2. The base condition, representing the grand mean across all group types, with no membership claims and all continuous variables at their mean values received 1.67 responses. Request specificity had a strong impact on community responsiveness: being specific increased reply counts by 55% ($p < .01$), but making vague, open-ended requests was no different than making no request at all. A likelihood-ratio test indicates that adding interactions between request type and group type does not improve the model, $\chi^2(6, N = 297) = 10.40$ $p > .10$, and so interactions are not included.

Table 4. Base Message With Specific and Open-Ended Request Manipulations

Request Type	Message Text
None	<p>Original (excerpt) from rec.food.cooking Subject: Chicken "paws" (speaking of making stock)</p> <p>Could not believe my eyes . . . chicken feet this morning at Wal-mart! :-) They were in the section with sweetbreads, tripe, heart, tongue and the other more unusual offerings. They were labeled "Chicken paws". <lol>Now that the weather is cooler, it's about time for some good chicken soup and <u>nothing</u> imho makes better chicken soup than using feet! Soooo rich and yummy when you are done.</p>
Specific	<p>Could not believe my eyes . . . I found chicken feet this morning at Wal-mart! :-) They were in the section with sweetbreads, tripe, heart, tongue and the other more unusual offerings. They were labeled "Chicken paws". <lol></p> <p>Now that the weather is cooler, it's about time for some good chicken soup and <u>nothing</u> imho makes better chicken soup than using feet! Soooo rich and yummy when you are done. <i>Are there any other good recipes using chicken paws?</i></p>
Open-ended	<p>Could not believe my eyes . . . I found chicken feet this morning at Wal-mart! :-) They were in the section with sweetbreads, tripe, heart, tongue and the other more unusual offerings. They were labeled "Chicken paws". <lol></p> <p>Now that the weather is cooler, it's about time for some good chicken soup and <u>nothing</u> imho makes better chicken soup than using feet! Soooo rich and yummy when you are done. <i>Wondering whether others have used chicken paws.</i></p>

General Discussion

The present research demonstrates the effectiveness of three proactive socialization strategies used by newcomers to online groups: (a) group-oriented membership claims, in which newcomers describe past participation in the group; (b) identity-oriented membership claims, in which they describe their similarity to the group's social category; and (c) information requests, in which they ask for help from the group. We found that these socialization strategies increase a group's likelihood of granting provisional membership, measured here by responding to the newcomers' messages, and that these membership grants, in turn, increase the likelihood that newcomers will continue participating in the group. The evidence is strongest for group-oriented membership

Table 5. Number of Replies as a Function of Specific and Open-Ended Requests

	Incidence Report Ratio	Standard Error	Expected No. of Replies
Intercept ^a			1.67
Has specific request	1.55**	0.27	2.56
Has open-ended request	0.94	0.16	1.57
Controls			
Political issue group ^b	.54***	0.07	0.90
Health support group ^b	1.21	0.15	2.02
Technical group ^b	1.12	0.14	1.87
Group traffic that month; Ln(messages)	1.64***	0.10	—
Original reply count; Ln(messages)	1.19***	0.04	—

Note: All variables are binary unless otherwise indicated. Continuous variables were logged and centered. $N = 297$ messages.

a. The intercept represents the grand mean across all group types, with no requests and all continuous variables at their mean values.

b. Discussion group categories were converted to binary effect coding variables, with hobby groups omitted.

* $p < .05$. ** $p < .01$. *** $p < .001$.

claims and specific requests. Results are mixed for identity-oriented membership claims; correlational data showed that they were followed by increased membership grants, but experimental evidence did not.

For group-oriented membership claims, both correlational and experimental evidence across a wide variety of online discussion groups show that de-lurking statements lead to membership grants. Despite the negative connotations of the term *lurking*, groups respond positively when newcomers introduce themselves by asserting they have silently observed the group. By doing so, they reveal their prior investment in the group and a desire to learn the group's norms before demanding any of its attention. These displays presented newcomers as more desirable future full members, leading the community to grant them some of the privileges of membership, including a willingness to reply to them, answer their questions, and engage them in conversation.

The research also demonstrated that when newcomers make specific requests for information, they are more likely to receive the attention of other group members. Correlational evidence shows that any form of request increased response rates by approximately 40%, and experimental evidence shows that the effect is largely due to specific requests, which increase response rates by 55%. Though politeness theory would suggest that because newcomers have low social status, they should use indirect requests to save face (Brown &

Levinson, 1987), indirect requests were not as effective as messages that stated clearly and explicitly the information sought. Direct requests reduce burden by telling potential responders how to be helpful (Francik & Clark, 1985; Linde, 1998).

However, this study provides mixed evidence for the effectiveness of identity-based membership claims. In the correlational study, claims of membership in the social category around which the group is organized were associated with an approximately 40% increase in reply rates, but the experimental study failed to demonstrate that identity-based membership claims caused the increase. There are several possible explanations for this discrepancy: First, it is difficult to create good identity-based claims for 93 topics. The identity-based membership claims constructed for Study 2 were also shorter and more generic than those found in the wild. Previous research suggests that longer messages are less likely to receive replies (Whittaker, Terveen, Hill, & Cherny, 2003), so these claims were intentionally brief and always placed at the beginning of messages. Also, they were designed both to generalize easily to other groups and to lend themselves to machine learning. The short, generic identity-based membership claims in Study 2 may have been too artificial to have the same magnitude of impact as those observed in Study 1. It is also possible that the machine learning technique used to measure identity-based membership claims in Study 1 may have actually measured something else, such as self-confidence, which would also be associated with first-person pronouns (Pennebaker, Mehl, & Niederhoffer, 2003). These findings suggest the need for future research into the effectiveness of different forms of self-disclosure in communities of strangers.

Implications for Socialization Theory

Although previous theory (e.g., Levine & Moreland, 1994; Van Maanen & Schein, 1979) has argued and empirical research (Bartel, et al., 2007; Bauer & Green, 1998) has demonstrated that socialization to groups and organizations is a bidirectional process in which newcomers play a proactive role, the current research demonstrates that the way newcomers execute their moves matters. There are subtle differences in the ways newcomers attempt to capture a group's attention, and the language they use is important. For example, we find that linguistic specificity matters when requesting information. Though we only examine some forms of membership claims, there are others—such as helping others in the group or showing evidence of adhering to group norms—which may have similar effects of increasing membership grants.

By examining the details of the proactive steps newcomers take in their own socialization, this research helps overcome the biases in the socialization literature, which often focuses on the organization's actions (Van Maanen & Schein, 1979) or treats newcomers' behavior as generic feedback- or information-seeking (Ashford, Sluss, & Harrison, 2007; Bauer et al., 2007). Miller and Jablin (1991) propose that newcomers assess the "extent of fitting into the social environment" (p. 99). Just as newcomers to an organizational role might adopt a provisional self to try out a way of enacting the role and testing their fit by receiving feedback from others, so too do newcomers to groups seem to assert their membership as a probe and then get feedback, which in turn influences their likelihood of remaining in the group. Our research has examined the conversational move by which they execute this probe-and-interpret process.

This work also demonstrates the usefulness of studying socialization processes in online groups. Although online groups differ in some important ways from conventional groups and organizations, they socialize newcomers in ways that are analogous to those used in conventional groups. Because interactions in many online groups are logged, researchers can gain insights into the details of socialization processes that would be difficult to observe in conventional groups and extend the methods for studying group and organizational socialization beyond the retrospective interviews and surveys that have dominated research in this area. Although the current research has examined the ways in which newcomers present themselves in conversationally oriented groups, other research has examined socialization processes from both the newcomer and group's perspective in online production groups as well as conversational ones (e.g., Ducheneaut, 2006; Von Krogh et al., 2003; Wang et al., 2008). Similar results are emerging in other online domains. For example, newcomers who join project groups in the online encyclopedia Wikipedia participate more actively and edit project articles more if old-timer project members respond to them in the first weeks after they join (Choi, Alexander, Kraut, & Levine, in press).

Practical Implications for Online Groups

This research can help inform the design of online groups to foster more successful socialization. Online groups could include instructional materials, such as FAQs or templates to help newcomers learn how to engage the group based on the successful socialization moves identified here. In addition, the machine learning models from Study 1 could also form the basis of socialization bots, or automatic agents embedded in e-mail or newsgroup software,

analogous to a grammar checker, that diagnose newcomers' attempts to join groups and intervene at the point of writing, suggesting wording to enhance the community's likeliness to respond.

One concern with training everyone in online groups to use the *right* socialization moves is scalability. An automatic socialization agent might result in a flood of indistinguishable messages filled with "I've been lurking" claims or very specific questions. However, increasing the quantity of messages to the group does not necessitate decreasing the overall quality of the group. Successful socialization into an online community requires learning both norms and politics (Ducheneaut, 2006), so an assistant that improves rhetoric allows the author to focus on strengthening the message substance and its ability to attract allies. Allowing novice members to learn group norms faster and have successful interactions earlier may increase the pool of available responders to future messages.

Furthermore, this research demonstrates the usefulness of using machine learning techniques for studying social psychological processes in online groups. Simple text analysis tools, whether guided a priori by theory or used in a more exploratory way, allow researchers to discover linguistic patterns correlated with psychological outcomes of interest in data sets too large to code manually. When used in an exploratory way and combined with sensitivity analysis, these automated tools may unveil social patterns that are robust—particularly those that are surprising—that may indicate community phenomena needing further research.

Limitations

One primary limitation of this work is that it is based only on observable behavior, and provides no direct evidence of the newcomers' or groups' intentions or other internal states. Based on these data, we cannot confirm that newcomers intended their rhetorical strategies to be a membership claim, or the communities' members intended their replies to be a membership grant. An alternative view is that these conversation strategies are not evidence of proactive socialization, but rather indications that the newcomers view the community as a repository of knowledge that they intend to exploit for a one-time informational need. Thus, they may be maximizing rhetorical strategies that are likely to get other people to respond without the intention to make persistent social connections with the group. This may be the case for some individuals in some groups, particularly in technical Q&A forums, where there is extremely high turnover. Yet previous studies of online groups show that a significant minority of newcomers continues active participation

(Arguello et al., 2006), and the roles of question asker and answerer are mutually dependent and constantly shifting (Tichon & Shapiro, 2003). Messages phrased similar to impersonal database queries are typically ignored (Galegher et al., 1998), and online discussion groups are more than simple information repositories (Wang et al., 2008). Thus, we focus on these linguistic patterns and the community's response as evidence of proactive socialization strategies.

Although we have asserted that parallels exist between socialization in conventional groups and online groups, online discussion groups are qualitatively different organizations than those typically studied by organizational behavior researchers, such as the workplace, and the implications of these differences need to be explored further. The online groups in the present research were comprised of large numbers of volunteers who have never met face-to-face, and thus legitimacy may be a greater issue to newcomers here than in a physical organization with a shared corporate identity. Yet the present findings give insight into the general process of newcomer socialization, and specifically the newcomers' conversational strategies, from data rarely available in the workplace. The high turnover within these groups is another difference, but largely reflects first-time posters who never returned, and thus gives us greater insight into the difference between successful and unsuccessful attempts to engage the organization. Much like a large, international corporation, in which the majority of applicants do not pass the informational interview stage, these groups reveal the concrete ways in which successful and unsuccessful newcomers attempt to shape their experience and how the organization responds to their attempts. Group norms persist in the small contingent of active members who remain month-to-month and in written form, in the publically accessible message archives, another source of data accessible to prospective members not available in conventional organizations.

In the present research, socialization from the group's perspective was operationalized with a simple metric, the number of responses members of the group gave to a newcomer. Though this metric can quickly be computed and would generalize to other kinds of conversational-based groups, the present study does not take into account the valence of the responses. Certainly a response of "you're not welcome here" would indicate unsuccessful socialization on the part of the newcomer. However, informal review of responses suggests that responses in technical, hobby, and support groups are overwhelmingly positive and indicate acceptance of the newcomer, suggesting that groups simply withhold responses from newcomers who use unsuccessful probes. Political issue groups, on the other hand, often have a norm of antagonistic debate on ideological issues, and so even unfriendly responses

from the group may indicate that the newcomer has successfully become a member and may continue getting the lively argument he or she hoped for. Future study would indicate closer analysis of message valence and its impact on a newcomer's initiation into a group.

Finally, only one type of technology infrastructure—online discussion groups—was studied. Other technological platforms might result in different response patterns, such as those online groups that require membership enrollment and secure access or others that rely on synchronous chat or blogs. Furthermore, within discussion groups, we focus only on newcomers' thread-starting messages. More research is needed to determine whether newcomers should make their first interaction with the group by starting new threads, or whether they should make a more gradual entrance by replying to existing threads first. In the Freenet project, for example, potential developers were unlikely to be successful proposing entirely new modules if they had not first garnered community attention by posting smaller bug fixes (Von Krogh et al., 2003). Automatic interventions for socialization strategies would then account for the message's position in the thread.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research and/or authorship of this article:

National Science Foundation (Grant Nos. IIS-0325049 and IIS-0729286, all authors received partial support from these grants). The interests of the National Science Foundation are not discussed or alluded to within the contribution.

Note

1. When we added requests to the membership claim manipulations in a sample of 82 messages, none of the manipulations showed any effect. This discrepancy suggests that the addition of two manipulations to a base message made the message too artificial. To deal with this problem, we report here results based on messages in which only membership claims were manipulated. We conducted an additional experiment manipulating requests and report this in Study 3.

References

- Ahuja, M. K., & Galvin, J. E. (2003). Socialization in virtual groups. *Journal of Management*, 29, 161-185.

- Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman & Hall.
- Altman, I., & Taylor, D. (1973). *Social penetration: The development of interpersonal relationships*. New York: Holt.
- Anderson, C. M., Riddle, B. L., & Martin, M. (1999). Socialization processes in groups. In L. R. Frey, D. S. Gouran, & M. S. Poole (Eds.), *The handbook of group communication theory and research* (pp. 139-166). Thousand Oaks, CA: Sage.
- Arguello, J., Butler, B., Joyce, E., Kraut, R., Ling, K., & Wang, X. (2006). Talk to me: Foundations for successful individual-group interactions in online communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 959-968). New York: ACM Press.
- Ashford, B., Sluss, D., & Harrison, S. (2007). Socialization in organizational contexts. In G. Hodgkinson & J. Ford (Eds.), *International review of industrial and organizational psychology* (pp. 1-70). New York: Wiley.
- Barge, J. K., & Schlueter, D. W. (2004). Memorable messages and newcomer socialization. *Western Journal of Communication*, 68, 233-257.
- Bartel, C. A., & Dutton, J. E. (2001). Ambiguous organizational memberships: Constructing organizational identities in interactions with others. In M. A. Hogg & D. Terry (Eds.), *Social identity processes in organizational contexts* (pp. 115-130). Philadelphia: Psychology Press.
- Bartel, C. A., Wrzesniewski, A., & Wiesenfeld, B. (2007). The struggle to establish organizational membership: Newcomer socialization in remote work contexts. In C. A. Bartel, S. Blader, & A. Wrzesniewski (Eds.), *Identity and the modern organization* (pp. 119-133). Mahwah, NJ: Lawrence Erlbaum.
- Bauer, T., Bodner T., Erdogan B., Truxillo D., & Tucker J. (2007). Newcomer adjustment during organizational socialization: A meta-analytic review of antecedents, outcomes, and methods. *Journal of Applied Psychology*, 92, 707-721.
- Bauer, T. N., & Green, S. G. (1998). Testing the combined effects of newcomer information seeking and manager behavior on socialization. *Journal of Applied Psychology*, 83, 72-83.
- Baym, N. (1993). Interpreting soap operas and creating community: Inside a computer-mediated fan culture. *Journal of Folklore Research*, 30, 143-176.
- Branscombe, N., Spears, R., Ellemers, N., & Doosje, B. (2002). Intragroup and intergroup evaluation effects on group behavior. *Personality and Social Psychology Bulletin*, 28, 744-753.
- Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge, UK: Cambridge University Press.
- Bryant, S. L., Forte, A., & Bruckman, A. (2005). Becoming Wikipedian: Transformation of participation in a collaborative online encyclopedia. In *Proceedings of*

- the 2005 International ACM SIGGROUP Conference on Supporting Group Work* (pp. 1-10). New York: ACM Press.
- Butler, B. (2001). Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research, 12*, 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.
- Carvalho, V., & Cohen, W. (2004). Learning to extract signature and reply lines from email. In *Proceedings of the Conference on Email and Anti-Spam*. Retrieved October 4, 2009, from <http://www.ceas.cc/papers-2004/135.pdf>
- Choi, B. R., Alexander, K., Kraut, R. E., & Levine, J. M. (in press). Socialization tactics in Wikipedia and their effects. In *Proceedings of the 2009 ACM Conference on Computer-Supported Cooperative Work*. New York: ACM Press.
- Clary, E. G., Snyder, M., Ridge, R. D., Copeland, J., Stukas, A. A., Haugen, J., & Miene, P. (1998). Understanding and assessing the motivations of volunteers: A functional approach. *Journal of Personality and Social Psychology, 74*, 1516-1530.
- Cohen, W. W. (2004). *Minorthird: Methods for identifying names and ontological relations in text using heuristics for inducing regularities from data*. Retrieved October 1, 2007, from <http://minorthird.sourceforge.net>
- Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science, 15*, 687-693.
- Collins, N., & Miller, L. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin, 116*, 457-475.
- Cramton, C. (2002). Attribution in distributed work groups. In P. Hinds & S. Kiesler (Eds.), *Distributed work* (pp. 191-212). Cambridge: MIT Press.
- Dönmez, P., Rosé, C., Stegmann, K., Weinberger, A., & Fischer, F. (2005). Supporting CSDL with automatic corpus analysis technology. In T. Koschmann, D. Suthers, & T. W. Chan (Eds.), *Proceedings of the International Conference on Computer Support for Collaborative Learning* (pp. 125-134). Hillsdale, NJ: Lawrence Erlbaum.
- Ducheneaut, N. (2006). Socialization in an open source software community: A socio-technical analysis. *Computer Supported Cooperative Work, 14*, 323-368.
- Facebook. (2009). *Press room: Statistics*. Retrieved August 28, 2009, from <http://www.facebook.com/press/info.php?statistics>
- Fisher, D., Smith, M., & Welser, H. (2006). You are who you talk to: Detecting roles in Usenet newsgroups. In *Proceedings of the 39th Hawaii International Conference on System Sciences* (HICSS '06; Vol. 3, p. 59b). Retrieved October 5, 2009, from <http://www2.computer.org/portal/web/csdl/proceedings/h#4>

- Flanagin, A. J., & Waldeck, J. H. (2004). Technology use and organizational newcomer socialization. *Journal of Business Communication, 41*, 137-165.
- Francik, E., & Clark, H. (1985). How to make requests that overcome obstacles to compliance. *Journal of Memory and Language, 24*, 560-568.
- Galegher, J., Sproull, L., & Kiesler, S. (1998). Legitimacy, authority, and community in electronic support groups. *Written Communication, 15*, 493-530.
- Griffin, A., Colella, A., & Goparaju, S. (2002). Newcomer and organizational socialization tactics: An interactionist perspective. *Human Resource Management Review, 10*, 453-474.
- Goffman, E. (1959). *The presentation of self in everyday life*. Garden City, NY: Doubleday.
- Hogg, M. A., & Terry, D. J. (2000). Social identity and self-categorization processes in organizational context. *Academy of Management Review, 25*, 121-140.
- Jablin, F. M. (2001). Organizational entry, assimilation, and disengagement/exit. In F. M. Jablin & L. L. Putnam (Eds.), *The new handbook of organizational communication: Advances in theory, research, and methods* (pp. 732-818). Thousand Oaks, CA: Sage.
- Jones, G. R. (1986). Socialization tactics, self-efficacy, and newcomers' adjustments to organizations. *Academy of Management Journal, 29*, 262-279.
- Jones, Q., Rafaeli, S., & Ravid, G. (2004). Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research, 15*, 194-210.
- Joyce, E., & Kraut, R. (2006). Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communication, 11*(3). Retrieved September 22, 2009 from <http://jcmc.indiana.edu/vol11/issue3/joyce.html>
- Lampe, C., & Johnston, E. (2005). Follow the (slash) dot: Effects of feedback on new members in an online community. In *Proceedings of the ACM SIGGROUP Conference on Supporting Group Work* (pp. 11-20). New York: ACM Press.
- Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.
- Lave, J., & Wenger, E. (1990). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Levine, J., & Moreland, R. (1994). Group socialization: Theory and research. *European Review of Social Psychology, 5*, 283-308.
- Levine, J. M., Moreland, R. L., & Ryan, C. (1998). Group socialization and intergroup relations. In C. Sedikides, J. Schopler, & C. Insko (Eds.), *Intergroup cognition and intergroup behavior* (pp. 283-308). Mahwah, NJ: Lawrence Erlbaum.
- Linde, C. (1998). The quantitative study of communicative success: Politeness and accidents in aviation discourse. *Language in Society, 17*, 375-399.
- McKenna, K., Green, A., & Gleason, M. (2002). Relationship formation on the Internet: What's the big attraction? *Journal of Social Issues, 58*, 9-31.

- Mignerey, J. T., Rubin, R. B., & Gorden, W. I. (1995). Organizational entry: An investigation of newcomer communication behavior and uncertainty. *Communication Research, 22*, 54-85.
- Miller, V., & Jablin, F. (1991). Information seeking during organizational entry: Influences, tactics, and a model of the process. *Academy of Management Review, 16*, 92-120.
- Moreland, R. L., & Levine, J. M. (1982). Socialization in small groups: Temporal changes in individual-group relations. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 15, pp. 137-192). New York: Academic Press.
- Morrison, E. W. (1993). Longitudinal study of the effects of information seeking on newcomer socialization. *Journal of Applied Psychology, 78*, 173-183.
- Morrison, E. W. (2002). Information seeking within organizations. *Human Communication Research, 28*, 229-242.
- Mortensen, M., & Hinds, P. (2002). Fuzzy teams: Boundary disagreement in distributed and collocated teams. In P. Hinds & S. Kiesler (Eds.), *Distributed work* (pp. 284-308). Cambridge: MIT Press.
- Myers, K. K., & Oetzel, J. G. (2003). Exploring the dimensions of organizational assimilation: Creating and validating a measure. *Communication Quarterly, 51*, 438-458.
- Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes, 41*, 211-236.
- Newman, M., Pennebaker, J., Berry, D., & Richards, J. (2003). Lying words: Predicting deception from linguistic style. *Personality and Social Psychology Bulletin, 29*, 665-675.
- Noel, J., Wann, D., & Branscombe, N. (1995). Peripheral ingroup membership status and public negativity toward outgroups. *Journal of Personality and Social Psychology, 68*, 127-137.
- Nonnecke, B., & Preece, J. (2000). Lurker demographics: Counting the silent. In *Proceedings of the SIGCHI 2000 Conference on Human Factors in Computing Systems* (pp. 73-80). New York: ACM Press.
- Pennebaker, J., Mehl, M., & Niederhoffer, K. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*, 547-577.
- Preece, J., Nonnecke, B., & Andrews, D. (2004). The top five reasons for lurking: Improving community experiences for everyone. *Computers in Human Behavior, 20*, 201-223.
- Rafaeli, S., Ravid, G., & Soroka, V. (2004). De-lurking in virtual communities: A social communication network approach to measuring the effects of social and cultural capital. In *Proceedings of the 37th Annual Hawaii International Conference on Systems Sciences (HICSS'04; p. 70203)*. Big Island, HI: IEEE Press.

- Retrieved October 5, 2009, from <http://www2.computer.org/portal/web/csdl/abs/proceedings/hicss/2004/2056/07/2056toc.htm>
- Raymond, E., & Moen, R. (2006). *How to ask questions the smart way*. Retrieved February 6, 2008, from <http://www.catb.org/~esr/faqs/smart-questions.html>
- Reichers, A. E. (1987). An interactionist perspective on newcomer socialization rates. *Academy of Management Review*, *12*, 278-287.
- Ridings, C., & Gefen, D. (2004). Virtual community attraction: Why people hang out online. *Journal of Computer-Mediated Communication*, *10*(1). Retrieved September 22, 2009, from http://jcmc.indiana.edu/vol10/issue1/ridings_gefen.html
- Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*, 696-735.
- Saks, A. M., Uggerslev, K. L., & Fassina, N. E. (2007). Socialization tactics and newcomer adjustment: A meta-analytic review and test of a model. *Journal of Vocational Behavior*, *70*, 413-446.
- Smith, M. (2003). Measures and maps of usenet. In C. Leug & D. Fisher (Eds.), *From Usenet to CoWebs: Interacting with social information* (pp. 47-78). Berlin: Springer Verlag.
- Smith, M. (2004). *Netscan: A social accounting search engine*. Seattle, WA: Community Technologies Group, Microsoft Corporation. Retrieved June 21, 2007, from <http://netscan.research.microsoft.com>
- Tichon, J., & Shapiro, M. (2003). The process of sharing social support in cyberspace. *CyberPsychology & Behavior*, *6*, 161-170.
- Van Maanen, J., & Schein, E. (1979). Toward a theory of organizational socialization. In B. M. Staw (Ed.), *Research in organizational behavior* (Vol. 1, pp. 209-264). Oxford, UK: Elsevier.
- Von Krogh, G., Spaeth, S., Lakhani, K., & Hippel, E. (2003). Community, joining, and specialization in open source software innovation: A case study. *Research Policy*, *32*, 1217-1241.
- Wanberg, C., & Kammeyer-Mueller, J. (2000). Predictors and outcomes of proactivity in the socialization process. *Journal of Applied Psychology*, *85*, 373-385.
- Wang, X. (2007). *An ecological perspective on online communities*. Unpublished doctoral dissertation, University of Pittsburgh, PA. Retrieved September 22, 2009, from <http://etd.library.pitt.edu/ETD/available/etd-08302007-123310/unrestricted/xwang2007.pdf>
- Wang, X., Kraut, R., Butler, B., Joyce, E., & Burke, M. (2008). *A relationship model of online communities*. Unpublished manuscript.
- Wasko, M., & Faraj, S. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, *29*, 35-57.

Whittaker, S., Terveen, L., Hill, W., & Cherny, L., (2003). The dynamics of mass interaction. In C. Leug & D. Fisher (Eds.), *From Usenet to CoWebs: Interacting with social information* (pp. 79-91). Berlin: Springer Verlag.

Bios

Moira Burke is a PhD student in the Human-Computer Interaction Institute at Carnegie Mellon University. She studies social aspects of computer-mediated communication and online communities, including newcomer socialization, collective decision making, and feelings of well-being. She also focuses on social uses of the Internet by adults on the autism spectrum.

Robert Kraut is the Herbert A. Simon Professor of Human-Computer Interaction in the School of Computer Science and Tepper School of Business at Carnegie Mellon University. He has broad interests in the design and social impact of computing and has conducted empirical research on online communities, the social impact of the Internet, the design of information technology for small-group intellectual work, the communication needs of collaborating scientists, the impact of computer networks on organizations, office automation and employment quality, and technology and home-based employment.

Elisabeth Joyce is a professor in the Department of English and Theatre Arts at Edinboro University of Pennsylvania. Her research interests include 20th-century avant-garde poetry's affiliations with other disciplines and online community self-governance and commitment.