Carnegie Mellon University
Research Showcase @ CMU

Human-Computer Interaction Institute | School of Computer Science

2007

Introductions and Questions: Rhetorical Strategies that Elicit Response in Online Communities

Moira Burke
Carnegie Mellon University

Elisabeth Joyce
Edinboro University of Pennsylvania

Tackjin Kim
Carnegie Mellon University

Vivek Anand
Carnegie Mellon University

Robert Kraut
Carnegie Mellon University

Follow this and additional works at: http://repository.cmu.edu/hcii

Recommended Citation
.

This Conference Proceeding is brought to you for free and open access by the School of Computer Science at Research Showcase @ CMU. It has been accepted for inclusion in Human-Computer Interaction Institute by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

Introductions and Questions: Rhetorical Strategies That Elicit Response in Online Communities

Moira Burke[1], Elisabeth Joyce[2], Tackjin Kim[1], Vivek Anand[1], Robert Kraut[1]
[1]Carnegie Mellon University, Pittsburgh, PA 15232
[2]Edinboro University, Edinboro, PA 16444
{moira, ttk, vivekanand, robert.kraut}@cmu.edu, ejoyce@edinboro.edu

Abstract

Conversation is a critical element to the success of online communities, both for the communities as a whole and the individual members. Yet conversations often fail: Forty percent of potential thread-starting messages in Usenet groups receive no response. When this happens, those who are trying to get information, support, or discussion from a group receive no benefit from their efforts. This paper presents a multi-method series of studies into community responsiveness to two rhetorical strategies: self-disclosing introductions and making requests. Introductions signal legitimacy and commitment, and requests reduce community effort by making clear what is needed. An analysis of a longitudinal sample of 40,931 messages from 99 Usenet groups shows that introductions referencing lurking and a personal connection to the topic of discussion increase the likelihood of getting a reply. So does making requests. Machine learning detects introductions and requests automatically within the text. A pair of experiments in which introductions and requests are added or removed from previously posted messages and reposted to Usenet groups shows that introductions referencing the group history cause an increase in reply counts, but requests do not. The research will inform the design of an automatic intervention triggered at the time messages are written that makes suggestions for strengthening rhetorical strategies when the community is unlikely to respond.

1 Introduction

Online communities allow millions of people who would never meet in person to interact. People join web-based discussion boards, email lists, and chat rooms for friendship, social support, entertainment, and information on technical, health, and leisure activities [24]. And they do so in droves. One of the earliest networks of online communities, Usenet, had over nine million unique contributors, 250 million messages, and approximately 200,000 active groups in 2003 [27], while the newer MySpace, founded in 2003, attracts a quarter million new members every day [27].

1.1 Conversation Is Critical to Success

These millions of participants experience online communities through text. While sites may include photographs, music clips, and immersive virtual worlds, the majority of online communities are text-based [14]. People go online to talk (write) and listen (read), and so conversation is a critical factor in an online community's success [1].

In order to survive, online communities, like face-to-face groups, must meet the needs of individual members and the group as a whole [4,14,22], and conversation is the mechanism through which this occurs. It is through reading archived conversations that newcomers determine whether the potential benefit of membership is worth the cost of participation [4,17], and groups need a steady supply of newcomers to replenish membership levels over time. Current members solicit information and steer the group toward topics they care about by starting conversations. People start conversations hoping to derive benefit from the group; depending on the response they get members and prospective members will either continue to participate or they will leave. Therefore, the community's response to conversational overtures is particularly important.

Previous research has shown that a simple measure of community responsiveness—whether a poster gets a reply—is associated with an increased likelihood of posting again [12] and increased speed in posting again [16]. The effect is stronger for newcomers [1]. Entering a group causes a period of mutual assessment, in which the prospective member and the group evaluate the expected benefit of the other [17]; a reply to a newcomer signals acceptance in the group and leads to more committed behavior by the newcomer. Posters interpret silence, on the other hand, as rudeness or unfriendliness [6]. Baym found that praise-filled replies in a soap opera group were a motivating force for posters [2]. Replies signal that the community believes the author is a valuable member worth its attention, and the author reciprocates by writing more and replying to others. Conversations transition newcomers into committed contributors.

Yet problems with community responsiveness persist. Forty percent of potential thread-starting messages in Usenet go unanswered [29,12]. Although the non-response rate varies with community type and is lower in health support groups and some closed communities, in almost all online communities a substantial minority of attempts to start conversation result in failure.

Some messages are more likely than others to get a reply, and the difference is often in the wording. This paper describes a series of studies investigating rhetorical strategies—specifically introductions and requests—and community responsiveness. The first is a correlational analysis of approximately 41,000 messages from 99 Usenet groups, using machine learning to detect introductions and requests. This is followed by experiments in which introductions and requests in year-old messages were added or removed and reposted to the group to see if the presence of these features changed the community response.

2 Factors That Increase Likelihood of Reply

Previous research has identified characteristics of the author, group, and message text that affect whether the community will respond. Kraut and colleagues found that newcomers were less likely to get replies than people who had posted even once prior in the group [1]. Group norms affect reply rates; support groups, for example, are more responsive to newcomers than political discussion groups [8]. Cross-posting improves the odds, as does posting to high-traffic groups on busy days, a finding contrary to that of Jones and colleagues [11], who argue that information overload reduces the chance of a single message being noticed and acted upon. Finally, being on-topic and using less complicated language increases likelihood of reply [1,31].

Beyond these low-level characteristics are the rhetorical strategies of the author. Two prevalent strategies that affect reply are self-disclosing introductions and making requests.

2.1 Introductions

Unlike conversation in face-to-face groups, messages in online communities can be ignored or dismissed as illegitimate. There is a cost to reading and replying to messages, so posters must prove they deserve community effort. Thus, legitimacy is a critical issue in computer-mediated communication, where posters struggle not only to be heard, but also to appear worthy of the group's attention [10].

Introductions describe the author's relationship with the group. They come in two varieties: **Group introductions** mention previous behavior in the group, such as lurking or knowledge of previous posts or group members:

> *"I've been reading here a while and wanted to ask."*

> *"Can we please stop the Otis bashing. I am a lurker that reads the group on a regular basis. When I went to the messages today, most of them were Otis etc. bashing."*

Topic introductions reveal a personal connection to the topic of discussion:

> *"I was recently diagnosed with Epilepsy. I've had what I thought were 'panic attacks' for several years, mostly since the teen years, but it turns out they have been various types of seizures"* to alt.support.epilepsy

> *"I just picked up PSP today as an impulse buy, and the much-heralded Lumines. I had to have it as it is from the same minds that brought us Rez, one of my favorite 'experiences' of all time (it's more than a game, IMHO)."* to rec.games.video.sega

Introductions serve two key purposes: signaling legitimacy and signaling commitment. Introductions use appropriate in-group vocabulary, demonstrate personal expertise, and indirectly show the author's relationship to the group by showing his or her legitimate relationship to the topic around which the group organizes. Disclosure also signals a willingness to be vulnerable; an act

that fosters the building of close relationships [19]. Galegher and colleagues found that in online support groups, posters introduced themselves with personal histories, increasing their perceived legitimacy; requests that did not contain introductions came across as impersonal "database queries" and were unlikely to elicit response [10]. Online community design handbooks (e.g. [14]) promote the use of profiles—system-supported introductions—to allow members to get to know each other.

Referencing lurking, specifically, demonstrates that the author has committed time and effort in reading what others have already said. Though "lurking" has negative connotations in the offline world, it does not necessarily imply "free-loading" online [23]. These forms of introductions signal a desire to belong and demonstrate that the author is trying to learn more about the group and its norms before participating more actively. Baym found that "delurking" posts in a soap opera discussion group flagged the entrance of a new member into the community, and were often followed by "welcoming committee" posts [2].

Pronoun use in introductions is of particular interest, in that pronouns communicate the author's connection to the audience. Previous research suggests that first-person singular pronouns [e.g. I, me, my] indicate disclosure, first-person plural [we, us, our] indicate solidarity or social identity, and third-person pronouns [she, his, they] reference individuals about which the group collectively knows or distinguish in-group and out-group members [3,21]. On the other hand, second-person pronouns [you] suggest social distance. Therefore, pronouns implicitly strengthen or weaken the author's rhetoric. First and third person pronouns should elicit greater community response, while second person pronouns should reduce it.

2.2 Requests

A second rhetorical strategy linked to community responsiveness is making requests. Requests and questions are calls to action; they make clear what the poster hopes to get from the group. They reduce effort of response by telling people exactly how they can be helpful. Also, conversation norms of turn-taking encourage responding to questions with answers [26]. Requests are a common form of "seed" message in online support groups, and given the diversity and volume of messages online, the threads with question-answer pairs reduce chaos and ambiguity [10].

For the purpose of this research, questions (e.g. "*What can I expect from chemotherapy?*") and indirect requests not in the form of a question (e.g. *I'm wondering what to expect from chemotherapy.*") are treated as having a similar purpose: posters are seeking a response from the group. Yet the difference in language between explicit questions and implicit requests may affect whether the group will respond. While indirect language often signals politeness, explicit requests are more likely to elicit response than indirect or polite requests [18,9]. Linde found that attempts to change topic in face-to-face conversation were less likely to be continued if they contained mitigating language such as

"would," "could," "please," and "right" [9]. Messages initiating new threads are implicit attempts to change topics, and so explicit requests within these messages should elicit greater community response.

2.3 Limitations of Previous Work

Previous studies of responsiveness in online groups have made methodological tradeoffs between population size and depth of analysis. Studies of millions of messages afford only shallow analyses of content, based on metadata and bulk metrics like word count [31,11]. Studies that delve more deeply into the linguistic content of the messages generally rely on human coders to read messages. This approach is only practical with smaller corpora of a few thousand messages from a handful of groups [1,12]. The present study uses machine learning to overcome this tradeoff: A simple decision tree learner was trained to automatically detect introductions and requests—both explicit and implicit—in a large corpus of 41,000 messages.

Furthermore, previous investigations of introductions [1,10] derived examples solely from support groups, where the typical introduction includes references to diagnosis, therapy, and daily life. However, introductions in other kinds of groups, such as server administration, economics policy, or yoga, may take largely different forms and affect community responsiveness to different degrees. For example, von Krogh and colleagues found that introductions in which newcomers described their technical skills were insufficient for joining a technical community. Instead, the newcomers had to post code fixes [15]. The present study includes introductions from health support, political, hobby, and technical groups.

Finally, previous research of archival data has been largely correlational. Introductions and requests have been associated with an increased likelihood of reply, but have not been shown to cause it. For example, delurking introductions are likely to come from people who have read many messages in the group, and it may be their general familiarity with the group and not the use of an introduction that causes the group to respond. The present study establishes causation by experimentally manipulating messages in live communities.

3 Study 1: Correlational Analysis of 41,000 Usenet Messages

The first study expands the previous work of Kraut and colleagues [1], which had a corpus of 6,000 messages from 8 groups, to a much larger corpus from a wider array of groups. The goals were twofold: To use machine learning to automatically detect introductions and requests across many groups, and to determine the effect of those strategies on community responsiveness.

3.1 Data Collection

The sample was drawn from the structural data and message text of 99 Usenet newsgroups from June 2003 to February 2005. The data included the combination of individual and structural data provided by the Microsoft Netscan project [29] and the text of the posted messages, which were downloaded from a Usenet archive. The Netscan database provided structural information such as 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 to a message.

The 99 groups represent a wide variety of topics and populations, comprising four general categories: health support, technical, hobby, and political issue discussion. 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 missing text from one group left 99. All groups had a minimum of four posts per week during the study period.

A sample of 40,931 messages was selected from the 2,179,999 messages posted to the 99 groups during the focal period. The messages were selected to be first posts in their threads, and thus were potential conversation starters rather than replies to ongoing conversations. A maximum of 500 of these potential thread-starters was randomly selected from each of the 99 groups, such that the authors were distinct and the full message text was available. Some groups had fewer than 500 messages meeting these criteria, and cross-posted duplicate messages were removed, resulting in a sample of 40,931 messages. Slightly more than half of the messages (57%) received a reply.

3.2 Measures

Dependent Variable

Community responsiveness is measured by a dummy variable *GotReply*, reflecting whether a focal message received a reply (1) or not (0). Results are qualitatively the same when the number of replies is used.

Independent Variables: Rhetorical Strategies

The following section describes the key features of introductions and requests, followed by a description of the machine learning algorithm and dictionary package that were used to detect them.

An **introduction** is a self-disclosing post to a group that contains most of the following features: first-person pronouns (though sometimes describing the situation of a third party), the age of the poster, the acknowledgment that this message is a first-time post (e.g., *"I've been lurking here*" or "*let me start by…*"), and a description of the poster's relationship to the topic *("I was diag-*

nosed in 1994 with BC…." to alt.support.cancer.breast or "*I'm doing street stuff and need a cheap uni that's just a step up from a beginner*" to rec.sport.unicycling).

A **request** asks for something from the group. Not all requests are proper questions with question marks, so requests were based on more than punctuation. Other features include interrogatives and reversed subject and verb ("*Does anybody know how to play it on guitar*." "*Wondering the best way to dissolve chocolate, besides eating it*."), indirect requests ("*I want*," "*I'm looking for*"), and references to help ("*suggestions," "advice," "recommendations*").

Minorthird, a machine learning and text classification toolkit, was used to classify messages' rhetorical content. In prior research, Minorthird has been used to identify signature files, quotations, and speech acts in email [5]. Minorthird's Boosted Stump Learner was trained with 392 human-coded examples of introductions drawn from 101 Usenet groups. Human coders looked for the features described above. A comparison of the machine classification with the human-coded training set for a 10-fold cross-validation gave a recall of 0.75, precision of 0.83, and Kappa of 0.63. Kappa is an estimate of agreement corrected for chance, and a Kappa of 0.7 is generally considered as good as a human observer. A dummy variable, *Has Introduction*, is 1 if Minorthird classified the message as containing an introduction and 0 otherwise. Message terms ranked highly informative for introductions by Minorthird included [I, my, now, years, during, been, month]. Strongly negative features included [you, people, think, very, other, $].

For requests, a human-coded training set of 1011 messages with requests from two Usenet groups was used. Request-training messages came from fewer groups because they were determined by human judges to be highly similar across group types, unlike introductions. Machine classification of requests had recall of 0.76, precision of 0.78, and a Kappa of 0.61. A dummy variable, *Has Request*, is 1 if Minorthird classified the message as containing a request and 0 otherwise. Highly informative features Minorthird ranked as positive indicators of requests included [thanks, anyone, I, wondering, what, help, seem, ?]. Additional interrogatives (e.g. how/where/who) were most likely not included because their frequency reduced their information value. Negative indicators of requests included [we, you, their, f**k, see, !].

Pennebaker's Linguistic Inquiry and Word Count (LIWC) software [20] was used to count low-level words related to introductions and requests that may not have been detected by Minorthird. The analysis included the percentage of pronouns in the message body, the number of question marks in the subject line, and the percentage of sentences with question marks in the message body.

Controls

Factors previously determined to affect the likelihood of getting a reply were included as controls in this analysis. Control variables are listed in Table 1, and include the group type, traffic, and message complexity (information overload may prevent any single message, especially one that requires effort to understand, from receiving a response) [11]. The inclusion of introductions necessar-

ily adds length to the message body, and as longer messages are less likely to receive replies [11], message length was controlled. Spam messages are unlikely to include introductions or requests, but are also unlikely to receive a response, and so were controlled. SpamAssassin was used to generate a dummy variable, *Spam*, which equals 1 if the message is likely spam, and 0 if not.

3.3 Results and Discussion

To examine the effect of the rhetorical strategies of introductions and requests on community responsiveness, the analysis focuses on predicting the likelihood that a thread-initiating message will elicit a response. Table 1 presents the results of a logistic regression on the binary dependent variable, *Got Reply*. The table shows the effect of each of the explanatory variables on the likelihood that a focal message received a reply. Because messages within a single newsgroup are not independent of each other, the analysis was conducted using the cluster feature in Stata's probit procedure, which adjusts the standard error of the coefficients to account for non-independence of observations.

Table 1. Probability of getting a reply

				Model 1 Context		Model 2 Rhetorical Strategies	
		Mean	S.D.	Δ*Prob*[1]	SE	Δ*Prob*	SE
	Poster is newcomer (0/1)	.37	.48	-.04***	.01	-.05***	.01
Group[2]	Issue group	.23	.42	-.27***	.01	-.19***	.01
	Hobby group	.28	.45	-.05***	.01	-.03***	.01
	Support group	.22	.42	-.02*	.01	.00	.01
	Msg is cross-posted (0/1)	.25	.43	.07***	.01	.11***	.01
Overload	Msgs over 21 months	1.1e5	2.3e5	.04***	.00	.04***	.00
	Msgs per day in group	68.57	131.9	.02***	.00	.02***	.00
	Characters in subject	35.63	27.52	-.08***	.00	-.07***	.00
	Words in msg body	367.2	1221	-.02***	.00	-.03***	.00
	% Words > 5 letters in subject	31.58	26.86	.01***	.00	.01***	.00
	% Words > 5 letters in body	18.16	10.60	.01***	.00	.00	.00
	% Abbrs. in subject	.44	3.62	-.02***	.00	-.02***	.00
	% Abbrs. In body	1.11	3.57	-.08***	.00	-.04***	.00
	Words per sentence in body	4.88	558.3	.04***	.00	.03***	.00
	% Numbers in subject	.49	1.70	-.04***	.00	-.04***	.00
	% Emoticons in subject	.16	1.86	-.01	.01	-.01***	.01
	% Emoticons in body	.19	.91	-.03***	.01	.03***	.01
	Topicality[3]	6.65	7.35	.02***	.00	-.02***	.00
Spam	% $ signs in subject	.02	.22	-.12***	.02	-.06**	.02
	% ! in subject	.11	.75	-.01	.01	.00	.01
	% ! in body	.99	38.32	-.02***	.00	-.02***	.00
	% CAPS in subject	.13	.17	-.20***	.01	-.15***	.02
	Spam (0/1)	.67	.92	-.05***	.01	-.04***	.01

Table 1. (cont.).

Introduction	Has Introduction (0/1)	.74	.44		.07***	.01
	%1st person sing. in subject	1.04	4.89		.00	.00
	%1st person sing. in body	2.92	3.77		.08***	.00
	% 1st person pl. in subject	.25	2.15		.01	.01
	% 1st person pl. in body	.48	1.23		.00	.00
	% 2nd person in subject	.54	3.33		-.01*	.00
	% 2nd person in body	.89	2.04		-.04***	.00
	% 3rd person in subject	.21	1.88		-.02***	.01
	% 3rd person in body	1.10	1.92		.02***	.00
Request	Has Request (0/1)	.29	.46		.06***	.01
	Num. qmarks in subject	2.24	7.48		.03***	.00
	% qmarks in body	.63	1.72		.05***	.00
				R^2=.13	R^2=.17	

Notes: N=40,931. * p < .05, ** p < .01, *** p < .001
1. Δ *Prob* is the change in the probability of getting a reply (probit regression).
2. Technical group is the missing category.
3. Topicality was measured using average document frequency (ADF), the number of messages in the group in which the content words of the focal message appeared one or more times, divided by the number of words in the focal message and total messages in the group. As a pre-process, messages were stemmed and functional terms (i.e. the, this, is) were removed. A high ADF indicates that the language in the focal message is widely used in the group.

All continuous measures have been logged (base 2). So, Δ *Prob* for continuous measures indicates the effect of doubling the independent variable on the probability of getting a reply. Δ *Prob* for binary variables (0=False, 1=True) indicates the effect of changing the variable from false to true.

Overall, messages had a 57% chance of getting a reply and the rhetorical content had a strong impact on whether the community responded. Posts that included introductions detected by Minorthird were 7% more likely to receive a reply than those without. Over and above introductions identified by the machine-learning algorithm, first person singular pronouns (i.e. I, me, my) increased the likelihood of reply: Every doubling of these pronouns in the message body increased likelihood by 8%. Doubling the third person pronouns (i.e. he, she, theirs) in the message body increased it by 2%. Pronoun effects sometimes differed in the subject line: First-person singular pronouns in subject lines were not significant, while third person pronouns reduced likelihood of reply. The reduced real estate may afford different ways of speaking in subject lines. Adding pronouns to the model reduces the effect of introductions when modeled without pronouns, indicating that the final model is not weighting pronouns redundantly. Alternatively, if we recognize that the machine learning classification of introductions is imperfect, the increase in reply rates from pronouns may be due to introduction language missed by the classification.

Requests also increased the likelihood of reply. Messages with requests detected by Minorthird were 6% more likely to receive a reply than those without, consistent with the theory that requests invoke linguistic norms that elicit response [26]. Over and above the requests detected by Minorthird, which included both explicit questions and implicit requests, question marks in the message body and subject line each increased the likelihood of reply. Every doubling of question marks in the subject line increased reply rates by 3%, and every doubling of sentences ending in question marks in the message body increased reply rates by 5%.

Study 1 shows that requests and self-disclosing introductions have a significant impact on the likelihood that a community will respond to a conversational overture. After controlling for numerous low-level characteristics of a message and the environment in which it was posted—including newsgroup traffic, message complexity, and spam—the high-level rhetorical strategies of the author affect the probability of getting a response. However, while 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 the success of online communities, the data are correlational. Therefore, Study 2 was designed to test causality of introductions and requests.

Furthermore, introductions within this corpus really come in two varieties: group introductions referencing the virtual group (e.g. "*I've been lurking*") and topic introductions referencing the common topic:

> *"I've been trying to find a good milk substitute for a long time and have been experimenting with some soy or soy-containing powders"* in rec.food.vegetarian-cooking

> *"i have a heavyness or tightness of chest, also i get tightning of throat. this is fith week. could this be the ms hug. my ms center closed and my nuro isn't ms specialist, he didn'y do spinal mri just brain which showed no recient lesions. every test for heart lungs ect is negative*." in alt.support.mult-sclerosis

The two varieties of introductions observed in the wild are consistent with those described in the literature [10], and yet the relative impact of each kind is unknown. Finally, requests within Study 1 included both implicit and explicit forms, but the literature suggests explicit questions may be more powerful than implicit requests [18,9]. Study 2 experimentally manipulates messages in current Usenet groups to test causality, to disentangle the effects of the two kinds of introductions, and to test the impact of explicit questions.

4 Study 2: Introduction Manipulations

In Study 2, previously posted messages from 93 Usenet groups were experimentally manipulated, with introductions added or removed, and then reposted to the original groups to see the impact of the manipulations on community response. The high turnover rate—only 28% of the posters in one month remain the following month [30]—suggests that there is little likelihood of the original message author or readers observing the reposted message.

4.1 Method

Of the 99 groups from Study 1, 93 were still active in July 2006. From each of the 93 groups, two potential thread-starting messages—not replies to existing threads—that were at least one year old were randomly selected from the Google Groups site. Spam, FAQ reminders, outdated topics, and obvious flaming were excluded. Messages were selected such that one had an explicit question—defined below—and one did not.

Two independent judges coded the two messages for the presence of introductions. **Group introductions** indicate the author's previous lurking or posting behavior. **Topic introductions** indicate the author's relationship to the topic the group cares about. Topic introductions were coded using the same guidelines as those used to train the machine learning algorithm in Study 1, with an additional caveat: The introduction had to include information beyond the simple facts necessary to diagnose a problem. Therefore, "*I'm running RedHat 9.0 with MySQL 5.1*" and *"I used Redek on my [surf]board*" would not be coded as topic introductions because technical specifications alone do not describe the author's personal connection to the topic, while "*I've been running Windows servers for the last three years and I've been thinking about switching to Linux*" is a topic introduction.

The judges also coded explicit questions as a control variable. *Explicit questions* were defined as phrases ending with a question mark or, if they end with other punctuation, those that would be grammatically correct questions if they ended with a question mark (e.g. *"Has anyone heard of this [quilting] block.*" or "*Will it fit in a 172.*"). Rhetorical questions were coded as explicit if they addressed the group (e.g. "*Why can't everyone stop whining?*") but not if they were part of a narrative that did not address the group (e.g. "*Boy, was I surprised or what*?"). Cohen's Kappa for inter-rater reliability was 0.91 for topic introductions, 0.87 for group introductions, and 1.0 for explicit questions.

For each of the "base" messages, three versions were derived: (1) One without an introduction, (2) One with a group introduction, and (3) One with a topic introduction. One of the three versions was the original message. Table 2 shows three versions of a base message.

Table 2. Three versions of an example message with topic/group introductions added

Topic Intro	Group Intro	Message Text
0	0	**Original from alt.support.cerebral-palsy** **Subject: Neuromove** Anyone had any experience with this device? http://www.neuromove.com/
0	1	I've been reading here for the last month and am ready to jump in. Anyone had any experience with this device? http://www.neuromove.com/
1	0	My son has cerebral-palsy and I've been looking for options. Anyone had any experience with this device? http://www.neuromove.com/

The experiment employed a 4 (Group category: Support, Technical, Issue, Hobby) x 3 (Introduction: None, Group, Topic) design. The dependent variable was the number of replies received within one week. A preliminary analysis of over 200 million threads from the Netscan database shows 96% of thread-starting messages that receive a reply do so within 24 hours, so a weeklong window is adequate to count replies. One of the three versions of each base message was randomly selected to fill the experimental conditions, with 14 messages for each of the 12 combinations. The 168 messages were posted in counterbalanced order, 24 messages each day for one week.

Research involving interactions with online communities requires special ethical considerations, and so the experiment included several precautions to protect participants. To ensure that the activity would not noticeably affect the character of the group, the messages comprised a small percentage of the groups' traffic that week (median 4.1% across groups, mean 17.1%). To insure that privacy norms were not violated, the experiment included only large public groups with no registration and high turnover (mean 72%/month) so that the appearance of newcomers repeating ideas is normative, as are replies to them. 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). Original author names were changed and personal URLs were removed from signature blocks. Carnegie Mellon's Institutional Review Board determined that the potential benefits of this line of research outweighed the minimal risk.

4.2 Results and Discussion

The measure of community responsiveness in Study 2 was the number of replies. Roughly two-thirds (65%) of the messages received at least one reply, and the number of replies ranged from 0 to 55. Because the distribution of replies was non-normal, as expected, the square root of the number of replies was taken in the model described below. Therefore, the dependent variable ranges from 0 to 7.42.

Of the 168 messages posted, the community recognized 6 as reposts of previous messages. Those messages had 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. Four of the five groups were hobby groups (including one group that caught both messages), suggesting a lower turnover rate or more robust collective memory in hobby groups.

Table 3 shows a model of the number of replies as a function of its rhetorical content. In addition to the independent variables, *Group Introduction* and *Topic Introduction*, the model controls for the presence of an explicit question, the number of replies it originally received, which can be treated as a proxy for how interesting or compelling the message topic and style were, and the kind of group it was posted to (Hobby, Issue, Support, and Technical). The intercept represents a message in the default group, Technical, with all of the binary ex-

planatory variables set to 0 (i.e, a message with no group introduction, no topic introduction, and no explicit questions). A message of this type received an average of 0.42 replies (0.65^2=0.42). Adding a group introduction increased the number of replies by 92% to 0.81 ($(0.65+0.25)^2$=0.81). Adding topic introductions did not significantly affect the number of replies. The presence of an explicit question was correlated with an increase in replies by 68% to 0.71 ($(0.65+0.19)^2$=0.71), though causality of explicit questions was not tested in Study 2 (it was in Study 3). As expected, the number of replies the message originally received—how interesting it was—was significantly correlated with the number of replies after reposting. The difference between the four group categories was not significant. 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.

Study 2 showed that group introductions almost doubled reply counts, independent of message content and context. Topic introductions were not found to have a significant impact, which could be related to how difficult it is to create them for 93 topics. Group introductions, on the other hand, follow a standard format independent of group. Saying "I've been reading here for a while" is equally effective in economics and depression groups. Explicit questions were, once again, significantly correlated with more replies, so Study 3 manipulates both questions and introductions.

Table 3. Number of replies (sqrt) in Study 2

	Beta	S.E.
Intercept (Tech category, no introductions, no questions)	.65 ****	.12
Has group introduction	.25 *	.12
Has topic introduction	-.10	.12
Has explicit question	.19 *	.09
Original # of replies (sqrt)	.26 ****	.04
Hobby group	.18	.14
Issue group	-.25	.16
Support group	.17	.14

Note. Ordinary Least Squares (OLS) regression. Continuous variables have been normalized (sqrt). So, a message in the default category, Technical, with no introductions and no explicit questions would get $.65^2$=.42 replies.
* $p < .05$, **** $p < .0001$ N=168 R^2=0.24

5 Study 3: Introduction and Question Manipulations

Study 3 followed the same method as Study 2 with some exceptions. The five groups in which reposting was noticed by the community were excluded, as was one with an inactive moderator, leaving 87 groups. Six versions of each

base message were created, one for each combination of request and introduction (see Table 4). One of the six was the original message. As in Study 2, introductions were added to the beginning of the base message, but explicit questions were added where appropriate, either at the beginning or the end.

The experiment employed a 4 (Group category: Support, Technical, Issue, Hobby) x 3 (Introduction: None, Group, Topic) x 2 (Request: Explicit, Not Explicit) design. One of the six versions of each base message was randomly selected to fill the experimental conditions, with 7 messages for each of the 24 combinations. The 168 messages were posted in counterbalanced order, 24 messages each day for one week.

Table 4. Six versions of one base message. The original had an explicit question and a group testimonial.

Explicit Q.	Topic	Group	Message Text
1	0	1	**Original from alt.support.hepatitis-c** **Subject: Med-i-cal in California** Does anyone here know if Med-i-cal covers the treatment for HepC?? I have no insurance. Thanks for any information, I have already learned so much from your group.
0	0	1	I'm wondering if anyone here knows if Med-i-cal covers the treatment for HepC. I have no insurance. Thanks for any information, I have already learned so much from your group.
1	1	0	I was diagnosed with HepC years ago and have been holding off getting treatment. Does anyone here know if Med-i-cal covers the treatment for HepC?? I have no insurance. Thanks for any information.
0	1	0	I was diagnosed with HepC years ago and have been holding off getting treatment. I'm wondering if anyone here knows if Med-i-cal covers the treatment for HepC. I have no insurance. Thanks for any information.
1	0	0	Does anyone here know if Med-i-cal covers the treatment for HepC?? I have no insurance. ThSanks for any information.
0	0	0	I'm wondering if anyone here knows if Med-i-cal covers the treatment for HepC. I have no insurance. Thanks for any information.

5.1 Results and Discussion

Model 1 of Table 5 shows the number of replies as a function of the message's rhetorical content. The overall reply rate was approximately the same as Study 2; 65% of the messages received a reply, and the intercept received a mean of 0.40 replies (0.63^2=0.40). However, at first glace, neither group introductions nor explicit questions caused an increase in the reply count.

Yet there was no change in the way group introductions were manipulated in Studies 2 and 3. Therefore, we would expect them to cause an increase in replies, as they did in Study 2. This discrepancy suggests that the additional ma-

nipulation of questions may have interfered with the introduction manipulation. Question text was changed from the original message text in approximately half of the messages in Study 3; excluding these messages leaves a subset of 83 messages where only the introduction text was manipulated, similar to the conditions of Study 2. Even with the low power from such a small set of messages, group introductions nearly tripled the number of responses $(0.62+0.34)^2=0.96$, and the results were significant ($p < 0.05$) (see Model 2 of Table 5).

Table 5. Number of replies (sqrt) in Study 3.

	Model 1 Explicit questions		Model 2 Question manipulations removed	
	Beta	S.E.	Beta	S.E.
Intercept	.63****	.12	.62****	.14
Has group introduction	.01	.12	.34*	.15
Has topic introduction	.04	.12	-.14	.14
Has explicit question	.02	.08	-.08	.11
Original # of replies (sqrt)	.27****	.05	.28****	.06
Hobby group	.23	.15	-.30	.16
Issue group	-.26	.15	-.24	.18
Support group	.08	.15	.30	.18
	N=168 R^2=0.16		N=83 R^2=0.25	

Note: Ordinary Least Squares regression. Continuous variables have been normalized (sqrt). * $p < .05$, **** $p < .0001$

Furthermore, the results of Study 3 suggest that further investigation is needed to disentangle the effects of explicit and implicit requests. In the natural corpus of Study 1, explicit questions are correlated with an increase in replies, yet they do not cause it experimentally. The literature suggests explicit questions are more effective than implicit requests, yet that may not be the case here, or the manipulations need to be stronger. Lumping implicit requests in the non-explicit category may reduce its contrast with explicit questions.

Half of the messages in the 83-message subset of Study 3 had original, implicit requests. Collapsing the 83 messages with those of Study 2 reveals that the presence of any form of request—explicit or implicit—is correlated with an increase in replies from 0.35 to 0.59 $(0.59 + 0.18)^2=0.59$, an increase of 69% ($p < 0.05$). These correlational results suggest that requests—whether explicit or implicit—increase replies. However, the fact that adding explicit questions to messages that did not have them interfered with the effect of group introductions suggests that something about the request manipulation was not successful.

A more meaningful distinction may be between open-ended and specific requests. Netiquette pages such as Raymond and Moen's "How to Ask Questions the Smart Way" [24] have detailed guidelines for making requests effectively,

including being specific and explicit: "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."

Finally, topic introductions in Studies 2 and 3 were shorter and more generic than those found in the wild. Previous research suggests that longer messages are less likely to receive replies, so these introductions were intentionally brief and prepended to messages. Also, they were designed to easily generalize to other groups and lend themselves to machine learning. The short, generic topic introductions in Studies 2 and 3 may have been too artificial to have the same magnitude of impact as those in Study 1.

6 General Discussion

When a person posts a message to an online community, rhetoric matters. Including self-disclosing introductions demonstrates legitimacy and commitment to the group, while making requests makes it easier for repliers to help and implicitly calls the group to answer. The present studies show strong correlational and experimental evidence that introductions increase the likelihood that others will respond, and that the effect holds for numerous kinds of online groups. It also presents correlational but not experimental evidence that requests do the same.

This research can help inform the design of an intervention at the point messages are written, in the form of an agent embedded in email or newsgroup software, suggesting improved wording when the community is unlikely to respond.

But will it scale? Will this kind of automatic assistant result in a flood of indistinguishable messages to online communities? Or will it fuel an arms race of "better," more attention-getting rhetoric? This brings us to the issue of dual criteria for judging success in online communities: what is good for the individual may not be good for the group. Yet, by allowing novice members to learn group norms faster and have successful interactions earlier, we may increase the pool of available responders to future messages. Successful socialization into an online community requires learning both norms and politics [7], so an assistant that improves rhetoric allows the author to focus on strengthening the message substance and its ability to attract allies. So, increasing the quantity of messages to the group does not necessitate decreasing quality.

One final issue is 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 [15]. We intend to further analyze the Netscan data to determine whether Usenet group members have better long-term outcomes when they first appear in an ongoing thread rather than starting a new topic. Auto-

matic interventions for rhetorical strategies would then account for the message's position in the thread.

6.1 Limitations and Directions for Future Work

The present research assumes that all thread-initiating messages are seeking replies. This is likely true for the majority, given that people go online seeking things: information, friendship, and support [24], but there are certainly exceptions. We excluded FAQ reminders and spam from Studies 2 and 3 because the authors clearly do not intend to get replies. But the intention of other message types is less clear. Issue and health groups, for instance, have numerous thread stumps where members posted links to articles about regulations or drug tests. Are the authors just trying to be helpful, or do they want others to respond?

The machine learning model for introductions currently lumps both topic and group introductions together. Studies 2 and 3 indicate the need to develop separate models to determine their relative impact so that interventions based on introductions are most effective.

Finally, only one type of technology infrastructure, Usenet newsgroups, was studied. Other technologies might result in different response patterns. Further research can better understand the role of specific technologies in shaping responsiveness by studying groups that use technologies such as listservs, forums, or blogs.

7 Conclusion

Getting others in an online community to talk to you is difficult, but certain rhetorical strategies increase the odds. Introductions referencing the virtual group and topic, as well as requests, are correlated with an increased community response, over and above other contextual and linguistic factors, such as message length and newsgroup traffic. Group introductions like *"I've been lurking for a few months*" have been proven to double the number of replies. These results hold for a wide variety of populations, and so can be used to develop automatic interventions to improve messages before they are sent to the community, facilitating more successful interactions.

8 Acknowledgments

The authors would like to thank Xiaoqing Wang, Brian Butler, Kim Ling, Zoe Ouyang, and Kenneth Chan for feedback and assistance. This work was supported by NSF Grant IIS-0325049 and an AT&T Labs Research Fellowship.

9 References

1. Arguello, J., Butler, B., Joyce, E., Kraut, R., Ling, K., and Wang, X. Talk to me: Foundations for successful individual-group interactions in online communities. In *Proc. CHI 2006*, ACM Press (2006), 959-968.
2. Baym, N. Interpreting soap operas and creating community: Inside a computer-mediated fan culture. *Journal of Folklore Research*, 30 (1993), 143-176.
3. Brown, R., and Gilman, A. *The pronouns of power and solidarity*. Bobbs-Merril, 1960.
4. Butler, B. Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures. *Information Systems Research* 12, 4 (2001), 346-362.
5. Carvalho, V. and Cohen, W. Learning to Extract Signature and Reply Lines from Email. Paper presented at the Conference on Email and Anti-Spam. 2004.
6. Cramton, C. Attribution in distributed work groups. *Distributed Work*, Hinds, P. and Kiesler, S., Eds.. Cambridge, MA: MIT Press, 2002, 191-212.
7. Ducheneaut, N. Socialization in an open source software community: A socio-technical analysis. *Computer Supported Cooperative Work* 14 (2006), 323-368.
8. Fisher, D., Smith, M., & Welser, H. You are who you talk to: Detecting roles in Usenet newsgroups. Paper presented at the *39th Hawaii International Conference on System Sciences*, 2006.
9. Francik, E., and Clark, H. How to make requests that overcome obstacles to compliance. *Journal of Memory and Language,* 24 (1985), 560-568.
10. Galegher, J., Sproull, L., and Kiesler, S. Legitimacy, authority, and community in electronic support groups. *Written Communication* 15, 4 (1998), 493-530.
11. Jones, Q., Rafaeli, S., and Ravid, G. Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical *exploration*. *Information Systems Research*, 15, 2 (2004), 194-210.
12. Joyce, E. and Kraut, R. Predicting continued participation in newsgroups. *Journal of Computer Mediated Communication* 11, 3 (2006).
13. Karau, S. and K. Williams, Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality & Social Psychology* 65, 4(1993), 681-706.
14. Kim, A.., *Community Building on the Web: Secret Strategies for Successful Online Communities*. Berkeley, CA: Peachpit Press, 2000.
15. Krogh, G.v., Spaeth, S., Lakhani, K., and Hippel, E. Community, joining, and specialization in open source software innovation: A case study. *Research Policy*, 32 (2003), 1217-1241.
16. Lampe, C. and Johnston, E. Follow the (slash) dot: Effects of feedback on new members in an online community. In *Proc. GROUP 2005*, ACM Press (2005), 11-20.
17. Levine, J. and Moreland, R. Group socialization: Theory and research. *European Review of Social Psychology*. John Wiley & Sons: New York, 1994, 283-308.
18. Linde, C., The quantitative study of communicative success: Politeness and accidents in aviation discourse. *Language in Society*, 17, 1998, 375-399.
19. McKenna, K., Green, A., Gleason, M. Relationship formation on the Internet: What's the big attraction? *Journal of Social Issues* 58, 1 (2002), 9-31.
20. Pennebaker, J., Francis, M., and Booth, R. *Linguistic Inquiry and Word Count*. Mahwah, NJ: Erlbaum, 2001
21. Pennebaker, J., Mehl, M., and Niederhoffer, K. Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology* 54 (2003), 547-577.
22. Preece, J., Ed. *Online Communities: Designing Usability and Supporting Sociability*. John Wiley & Sons: New York, NY, 2000.
23. Preece, JK., Nonnecke, B., and Andrews, D. The top five reasons for lurking: Improving community experiences for everyone. *Computers in Human Behavior* 20, 1 (2004), 201-223.

24. Raymond, E, Moen, R. *How to Ask Questions the Smart Way* (2006). http://www.catb.org/~esr/faqs/smart-questions.html
25. Ridings, C. and Gefen, D. Virtual community attraction: Why people hang out online. *Journal of Computer-Mediated Communication*, 10, 1(2004).
26. Sacks, H., Jefferson, and Schegloff, E. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 4 (1974), 696-735.
27. Sellers, P. MySpace Cowboys. (2006). http://money.cnn.com/magazines/fortune/fortune_archive/2006/09/04/8384727
28. Smith, M. Measures and Maps of Usenet. *From Usenet to Cowebs: Interacting with Social Information*, C. Leug and D. Fisher, Editors. 2003, Springer Verlag: Amsterdam, Holland. p. 47-78.
29. Smith, M. Netscan: A Social Accounting Search Engine (http://netscan.research.microsoft.com). 2004, Community Technologies Group, Microsoft Corporation.
30. Wang, X., Butler, B., Joyce, E. An ecological perspective on online communities. In Academy of Management conference (2006).
31. Wasko, M., and Faraj, S. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. MIS Quarterly 29, 1 (2005), 35-57.
32. Whittaker, S., Terveen, L., Hill W., and Cherny, L., The Dynamics of Mass Interaction. From Usenet to CoWebs: Interacting with Social Information, C. Leug and D. Fisher, Editors. 2003, Springer Verlag: Amsterdam, Holland. p. 79-91.