Information Systems Research
Vol. 15, No. 2, June 2004, pp. 194–210
ISSN 1047-7047 | EISSN 1526-5536 | 04 | 1502 | 0194

DOI 10.1287/isre.1040.0023

Information Overload and the Message Dynamics of Online Interaction Spaces: A Theoretical Model and Empirical Exploration

Quentin Jones
Department of Information Systems, College of Computing Sciences, New Jersey Institute of Technology, University Heights, Newark, New Jersey 07102, qgjones@acm.org

Gilad Ravid, Sheizaf Rafaeli
Center for the Study of the Information Society, Graduate School of Business Administration, University of Haifa, Mt. Carmel, Haifa 31905, Israel {gilad@ravid.org, sheizaf@rafaeli.net}

Online spaces that enable shared public interpersonal communications are of significant social, organizational, and economic importance. In this paper, a theoretical model and associated unobtrusive method are proposed for researching the relationship between online spaces and the behavior they host. The model focuses on the collective impact that individual information-overload coping strategies have on the dynamics of open, interactive public online group discourse. Empirical research was undertaken to assess the validity of both the method and the model, based on the analysis of over 2.65 million postings to 600 Usenet newsgroups over a 6-month period. Our findings support the assertion that individual strategies for coping with "information overload" have an observable impact on large-scale online group discourse. Evidence was found for the hypotheses that: (1) users are more likely to respond to simpler messages in overloaded mass interaction; (2) users are more likely to end active participation as the overloading of mass interaction increases; and (3) users are more likely to generate simpler responses as the overloading of mass interaction grows.

The theoretical model outlined offers insight into aspects of computer-mediated communication tool usability, technology design, and provides a road map for future empirical research.

Key words: human-computer interaction; computer-mediated communication; computer supported cooperative work; virtual community; information overload; interaction coping strategies; message dynamics; online group discourse
History: Marshall Scott Poole, Associate Editor. This paper was received on March 7, 2002, and was with the authors 11 months for 3 revisions.

1. Introduction

The online spaces used for shared public interpersonal communications (enabled by technologies such as e-mail lists, Internet relay chat channels, and Usenet newsgroups) are of significant social, organizational, and economic importance (Rheingold 1993, Hagel and Armstrong 1997, Butler 2001). To date, a large proportion of studies on the behavior of users of such online spaces have been discussed in terms of "virtual community" (e.g., Cherny 1999, Jones 1997, Wellman and Gulia 1999). The focus on "community" has been typically associated with social or group theories. Consequently, less attention has been paid to the impact the nature of such virtual spaces has on user behavior, where shared public online interactions occur. However, it is in the design of these online spaces, rather than users' social networks, where information systems managers can often exert the greatest level of control. Therefore, there is a need for information systems researchers to examine the nature of the relationship between the virtual spaces used for shared public online behavior and the interactions (postings or behaviors) such systems contain. We address this need by proposing a theoretical model and associated method for examining the relationship between public online interaction spaces and their message dynamics.

To understand the impact that the spaces of public online interpersonal interactions have on user behavior, methodologies are needed that are neither culture specific nor time specific. Unfortunately, the relative novelty of this research question means that no analytical technique leaps to mind as a preferred or even obvious method of choice. The difficulty in choosing an appropriate method is perhaps exacerbated by the fact that analyses cannot rely solely on in-depth examinations of a small number of unique spaces and associated users' online behavior. This is because such research is likely to amount to an ad hoc exploration of individual systems whose shared characteristics, constraints, trade-offs, and metrics are unknown. This suggests that various forms of virtual ethnography, using techniques such as participant observation, are less appropriate. We also know that normative statistics cannot be used to predict the relationship between a technology and user-behavior in a deterministic fashion (Jones 1997). This is because social outcomes are determined by social context (Spears et al. 1992), not technology. Examples include: group norms and social learning (Schmitz and Fulk 1991), social identity (Lea and Spears 1991), the make-up of the community of users (Kling and Gerson 1977), and work group cohesiveness (Fulk et al. 1992). Therefore, the particular way in which a technology is used is not determined by the technology itself, but rather is dependent on its social context. Consequently, it will not be possible to accurately predict the particular form that online behavior will take without a detailed understanding of its social context. This in turn suggests that purely normative statistical approaches to describing behavior in public virtual interaction spaces may cloud understanding of the range of ways in which a communication technology is used. Further, the likely result of such analysis would be contradictory findings that fail to take into account the complexities of the social context. Similarly, it is difficult to design laboratory studies with ecological validity because good experimental design requires control for naturally occurring variations in social context, thus narrowing the range of social contexts explored.

The limitations of the standard methodological approaches to understanding the relationship between virtual interaction spaces and the behaviors such systems contain suggest that alternative techniques are needed. Ideally, these alternatives would help us to understand the impact of particular types of computer-mediated communication technologies on online behavior, independently of the narrow social contexts of use. To achieve this end, we propose a theoretical model and associated method for examining the relationship between public online interaction spaces and their message dynamics. The model focuses on the collective impact that individual information-overload coping strategies have on the dynamics of interactive public online group discourse. The proposed method for examining such impacts is large-scale field studies. Using such an approach, it is possible to move the emphasis away from notions of "community," and its attention to people and their relations, to the impact that virtual spaces have on online behavior. Following an outline of the theoretical model and associated method, we describe the empirical research undertaken to assess its validity. This empirical research is based on the analysis of 2.65 million messages posted to 600 Usenet newsgroups over a 6-month period.

2. Theoretical Model

In this section we outline a model of public virtual interaction dynamics based on an analysis of the collective impact of individuals' information-overload coping strategies. *Virtual publics* are symbolically delineated, computer-mediated spaces, such as e-mail lists, newsgroups, Internet relay chat (IRC) channels, that enable a potentially wide range of individuals to attend and contribute to a shared set of computer-mediated interpersonal interactions. They can therefore be considered as relatively transparent and open. We adopt here the term *virtual public* for two reasons: It is important that we distinguish between cybersociety (Jones 1995), virtual communities (Jones 1997), and open online interactions spaces (Jones and Rafaeli 2000a). Second, we need a simple label to describe online, shared, interpersonal interaction spaces, whose membership and existence are fairly open for both observation and user participation. Following the presentation of our proposed constraints model to virtual public message dynamics, we suggest a methodology for testing various associated hypotheses.

Historically, communication technologies have been enablers of only a limited range of social interactions. For example, in the ancient Middle East clay tablets assisted signal transmissions: They carried information and helped reduce the amount of required interpersonal interactions. Use of this technology helped create many great ancient civilizations. Clay tablets were also a constraining technology. They were not easily reproduced and required physical presence for message transfer. Clay tablet technology both enabled and constrained certain social outcomes through its particular form of asynchronous communication. Similarly, despite the technological advances and popular thinking about the limitlessness of cyberspace, we assume that each computer-mediated communication (CMC) technology acts to enable only a limited range of social interactions. Further, it stands to reason that the range of social interactions enabled and constrained by different CMC technologies will vary. It follows that appropriate measures of interactions occurring in different types of virtual publics should shed light on the extent to which they enable and constrain user interactions.

The question raised by the discussion of historical communication media is what are the equivalent constraints to public online group communication? Two constraints are discussed in the literature: critical mass and information overload. We review each and show how information overload can be used to explain and explore the message dynamics of virtual publics.

2.1. Human Limits to Shared Public Online Interpersonal Interactions

2.1.1. Critical Mass. Licklider (1968) suggested that CMC was constrained by the necessity for a "critical mass" or minimum number of people to be available for the solving of various problems. In their study of one of the earliest computerized conferencing systems, Hiltz and Turoff (1978) noticed that some discourse groups were simply "too-small," and lacked a "critical mass" to sustain interactions. Users of these small groups either stopped using the system or engaged in what they referred to as "electronic migration" to larger groups. Markus (1987) and Rafaeli and LaRose (1993) noted how critical mass could also be used to make predictions about adoption rates of communication technologies. Palme (1995) expanded on the work of Hiltz and Turoff and proposed different values of critical mass as appropriate for different activities. Palme also proposed a linear "communication response function" to explain the group size threshold for sustainable CMC. However, we know that the relationship between user population and user contributions cannot be described by a simple linear function because there is a decrease in the likelihood of individual user's participation in public discourse (Thorn and Connolly 1987, Whittaker 1996). Further complicating the issue, the term "critical mass" sometimes refers to the chances of gaining a response to a message (Palme 1995) and sometimes the chances of new material being generated (Hiltz and Turoff 1978). Perhaps most importantly here, various authors (Rafaeli and LaRose 1993, Ackerman and Palen 1996, Bradner et al. 1999) have shown that what constitutes critical mass is primarily determined by the social context of discourse, rather than technology itself. Therefore the need for "critical mass" is not a technology-mediated constraint to online behavior. This means that critical mass is not an obvious predictor of structural relationships between virtual publics and the message dynamics they host (which is the focus of this paper).

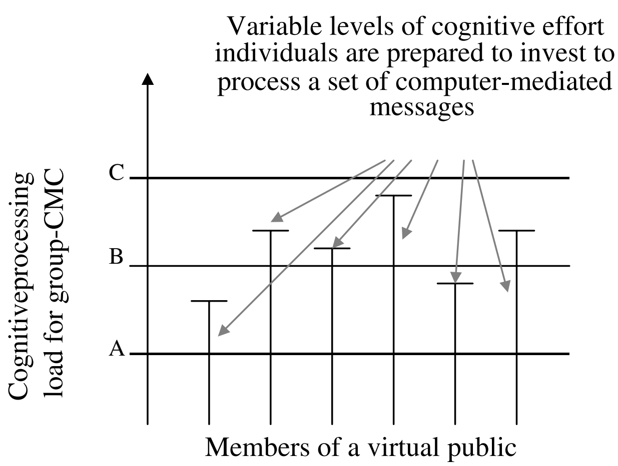
2.1.2. Information Overload. Information overload is the state of an individual (or system) in which not all communication inputs can be processed and utilized, leading to breakdown (Rogers and Agarwala-Rogers 1975). In the context of CMC, researchers refer to both *conversational overload*—when too many messages are delivered, so that individuals are unable to respond adequately (Whittaker et al. 1998)—and *information entropy* (Hiltz and Turoff 1985)—when incoming messages are not sufficiently organized to be easily recognized as significant or as part of a conversation's history.

Psychologists have long appreciated the limited capacity of people to store current information in memory (e.g., William James in the nineteenth century). The analysis of this information-overload producing limitation led in the 1950s to work in the field of cognitive psychology. Particularly influential in this regard were Miller's (1956) idea that we can process seven chunks of information (plus or minus two) and Broadbent's (1958) filter model of attention. In the physical or so called "real world," the maximum density and geographic spread of a culture's settlements are also linked to the management of information overload (Fletcher 1995).

For communities to function, individuals have to interact with each other. However, interaction involves the strain of dealing with other people, the effort of coping with the products of group activity such as noise and trash, and the effort we must expend to make communication possible. As a result, major changes in human settlement densities, such as those associated with the move from agrarianism to urbanism, have only been sustained historically when new assemblages of technological aids to interaction and communication have been developed, such as writing, clay tablets, and the telegraph. We reason that in a similar fashion to interactions in real settlements, sustainable interaction dynamics that occur using virtual publics (Jones and Rafaeli 2000a) are constrained by information overload. We claim that this is caused by limitations to an individual's ability to effectively process, over sustained periods, various patterns of virtual public interactions. *Communication load* is the processing effort required to deal with a set of messages or signals. An overly high communication load will result in information overload. Beyond a particular level of average-user communication processing, information overload will render the existing patterns of interactive public group communication unsustainable.

Figure 1 (from Jones and Rafaeli 2000a) illustrates how individual cognitive processing limits are linked to group communication. Three distinct cognitive load levels for groups CMC, A, B, and C are presented. At different cognitive load levels, different individuals are willing or able to process group messages. At Level A, every user can process the discourse. While at Level C, no one is able to process all the messages. Figure 1 illustrates that cognitive processing abilities of groups are not simply the sum of their individuals' cognitive processing capacities. Consequently, certain patterns of interactive group CMC cannot be sustained if the required processing effort (communication load) is higher than the maximum amount individuals can or are prepared to invest. The *average maximum communication load* (AvMaxCL) is therefore a crucial element in understanding virtual public message dynamics.

Figure 1 **Individual and Group Processing Limits**



Note. Reprinted with permission from Q. Jones and S. Rafaeli, What do virtual "Tells" tell? Placing cybersociety research into a hierarchy of social explanation. *Proc. 33rd Hawaii Internat. Conf. System Sci.*, Hawaii. © 2000, IEEE Press, Piscataway, NJ.

Individuals can take a range of actions to reduce the impact of information overload resulting from group CMC. Possible actions include: (1) increasing the effort, (2) learning new information management techniques, (3) failing to respond or attend to some messages, (4) producing simpler responses, (5) storing inputs and responding to them as time permits, (6) making more erroneous responses, and (7) ending participation in the group communication. These seven actions can be combined into two primary options for a population of experienced users who cannot easily adopt new information management techniques. The first option is simply to end participation. The second option is to change ongoing communicative behavior. We hypothesize that these seven individual responses to overloaded group CMC, can lead to observable impacts of virtual public interaction dynamics.

Although we do not have an exact measure of communication load, we do know that it relates to a number of message system characteristics. Users generally have to make more of an effort to reply coherently to a thread (a chain of messages) than to a single message (Lewis and Knowles 1997). Higher interactivity (see Rafaeli 1988 for a discussion of this term in regards to CMC) correlates with higher communication load. Similarly, a dense pattern of messages (high frequency of postings) will require quicker and more sustained processing by group members. Therefore, message density will also covary with communication load. It is also likely that an increase in interactional incoherence (such as fragmented sentences, agrammatical language, and interactionally disjointed messages) will also increase communication load (Herring 1999).

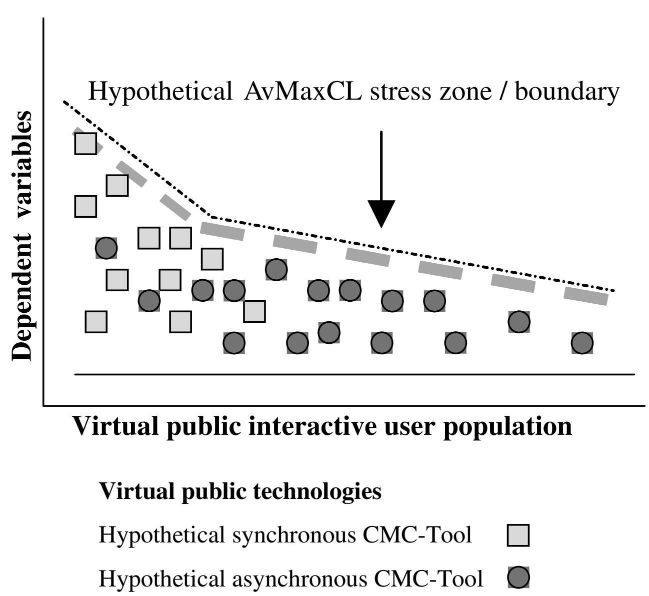
Abstracting the notion of individual information overload to impacts on virtual public interaction dynamics requires a systemic approach to virtual public discourse. This can be achieved by the recognition of virtual public discourse that is produced by members who are free to vocally participate, lurk, or unsubscribe as the output of a complex social system (Jones and Rafaeli 2000a; Rafaeli et al. 2004). Analysis of the systemic nature of social relations can be achieved through a focus on feedback loops (Forrester 1969). The logic is as follows, an increase in the membership of a virtual public is likely to be positively correlated with an increase in virtual public communication and communication load. However, individuals may not expand their involvement in virtual public communication indefinitely because of limited processing resources. Once virtual public communication becomes unmanageable or incoherent, involvement patterns will change. This in turn will impact subsequent discourse dynamics.

2.2. A Method for Understanding and Comparing the Message Dynamics of Virtual Publics

In this section we outline a method for empirically observing some of the impacts of the AvMaxCL on the interaction dynamics of virtual publics. The method is centered on the large-scale observations of *mass interaction*, the shared discourse between hundreds, thousands, or more individuals (Whittaker et al. 1998). When group CMC discourse approaches the AvMaxCL, a further increase in the communication load should result in an increased preference by individuals to engage in communicative actions that require less cognitive effort. This in turn will change the communication dynamics of the group CMC in question. For example, at AvMaxCL an increase in the number of interactive users may result in two of the seven individual responses to information overload listed above, shorter reply messages, and/or a higher turnover of users. In this case, both the proportion of users maintaining an active involvement and message length, can be explored as dependent variables as they are not constrained by the number of active users. Mass interaction, which in many cases is likely to be fairly overloaded, provides a unique opportunity to explore such relationships. This is because the large number of both messages and users involved in mass interaction should enable observations that may otherwise be hidden by differences between individuals and the social contexts of communication.

Figure 2 illustrates how the large-scale mapping of virtual public discourse should enable the modeling of the relationship between CMC technologies and communication load (Jones and Rafaeli 2000b). This figure suggests how a research methodology could be developed based on the analysis of virtual public mass interaction. The figure plots, for a number of virtual publics, an expected relationship between their user population and various free variables, such as the average level of message complexity. The figure shows that the stress zones caused by overloaded interactive communication can be empirically identified by mapping active participation in different virtual publics against various components of the communication load function. The plots of the different virtual publics vary widely because of

Figure 2 **Virtual Public Technologies and Message Processing Capacity**



Note. Reprinted with permission from Q. Jones and S. Rafaeli, What do virtual "Tells" tell? Placing cybersociety research into a hierarchy of social explanation. *Proc. 33rd Hawaii Internat. Conf. System Sci.*, Hawaii. © 2000, IEEE Press, Piscataway, NJ.

differences in their social context. The hypothetical synchronous virtual publics are located relatively close to the vertical-axis because they require user copresence for message exchange, which occurs fairly quickly and limits group size. On the other hand, the hypothetical plots of asynchronous virtual publics have larger user populations because users can take time to digest and respond to messages. Thus, the figure displays differences between virtual public technologies.

The approach outlined in Figure 2 does not assume that technology will determine use per se. Further, it does not suggest anything about the content of virtual public discourse within the boundaries imposed by technology, nor does it prescribe who will use one virtual public or another. To say something meaningful about the content of discourse, both social theory and a focus on context are required. It may also be the case that different types of social aims and social structures for online discourse (e.g., empathic as opposed to technical, moderated as opposed to unmoderated) may be associated with different stress boundaries. The model does not deal with this issue, although it provides a means for addressing it empirically. The model does not attempt to explain individual variations in the discourse patterns observed. Rather, it focuses on stress boundaries because we believe these provide the key link between technology and discourse structures.

In summary, our theory assumes that: (1) an individual must invest more cognitive resources to process large, complex group CMC than small-scale group CMC; (2) decisions made by individuals to employ various information-overload coping strategies will affect the dynamics of virtual public discourse; and (3) the nature of the cognitive resources required to process group CMC relates to the CMC technology in question. Building on these assumptions, we suggest a method for understanding the group-level usability of virtual publics. This method is field research, where the large-scale and naturally occurring patterns of sustained interactive online communication are mapped and analyzed. An aim of the method is to identify the collective impacts of individual information-overload coping strategies on virtual public interaction/discourse dynamics.

3. Hypotheses

The theory and method outlined above generate the following research questions.

(1) How does the volume of interactive group communication relate to the complexity of message content?

(2) How does the complexity of initial postings relate to the chances of gaining a response?

(3) How does the volume of interactive group communication relate to user participation patterns?

We address these research questions here through exploration of three hypothesized impacts of individual information overload on virtual public mass interaction dynamics in Usenet newsgroup postings. As discussed in further detail below, it is assumed here for all three hypotheses that a large proportion of Usenet newsgroups operate at AvMaxCL, and therefore we will be able to observe its impact by large-scale observation of Usenet interactions.

Hypothesis 1. *At AvMaxCL, as the number of interactive communications increases, there will be a decrease in the average complexity of response messages until asymptote.*

The communication load associated with the processing of group CMC can be weakly operationalized as the number of messages or interactive discussion threads per given time unit. Similarly message simplicity/complexity can be operationalized weakly in terms of various message characteristics, such as the number of words or new lines of text produced by the poster.

Hypothesis 2. *At AvMaxCL, simple group CMC messages will be more likely to generate responses than complex messages.*

We believe this will occur because people are likely to attend to messages that are hard to process. Similarly, they will also respond less to more complicated messages. We explore this empirically by examining if simpler messages are more likely to seed (start) new discussion threads than complex messages.

Hypothesis 3. *At AvMaxCL, as group CMC complexity increases, there will be an increased tendency for individuals to end or reduce active participation.*

We suggested above that disengagement is a strategy users will often adopt to cope with overloaded discourse. It follows then that the larger the number of individuals involved in overload group CMC discourse, the less stable the population of active participants.

4. Research Methodology

As noted above, field studies involving the mapping and analysis of large-scale, naturally occurring patterns of sustained interactive online communication are theoretically and methodologically appropriate. To implement such field studies and assess the hypotheses we have outlined, we need to choose appropriate virtual public technologies, collect large data samples, operationalize key concepts, and analyze virtual public interaction dynamics.

4.1. Usenet Technology

Virtual publics are supported by a wide variety of technologies, including the Usenet, Internet relay chat, e-mail lists, and avatar systems. The Usenet is a system of electronic bulletin boards, referred to as newsgroups. It is not a computer network, but rather a network of multilateral agreements among system administrators to cooperate on bulletin board management (Sproull and Faraj 1997). A number of practical considerations make Usenet newsgroups a useful virtual public technology with which to assess the first three hypotheses.

(1) The collection of hundreds to thousands of newsgroup user postings (interactions) is relatively straightforward. This is important because prior to this research it was not apparent how large a sample size was required to demonstrate the hypothesized phenomena.

(2) The capture of inter-user-interactivity data for analysis, such as discourse threading, for newsgroups is apparent (Liu 1999 outlines why this is not the case with most synchronous CMC).

(3) Anecdotal evidence suggests that a large percentage of Usenet newsgroups are overloaded (Smith 1999, Smith and Fiore 2001). The importance of this third feature of Usenet newsgroups is linked to the need to include a reasonable number of virtual public interactions operating at AvMaxCL in the sample.

4.2. Data Collection

Representative sampling of Usenet discourse is difficult; the Whittaker et al. (1998) solution was to produce a randomly stratified sample of English text-based Usenet newsgroups. They extracted 500 newsgroups from a subset of then-active, widely distributed newsgroups, which contained predominately English text-based conversational messages. For this project, we used the same core 500 newsgroups studied by Whittaker et al. (1998). An additional 100 newsgroups were also selected based on the Whittaker et al. approach. This allowed for 100 moderated groups to be selected. The complete content of 3,293,995 postings was collected over 8 months. From this larger sample, all of the 2,652,552 messages collected over the 6 months from August 1, 1999, to February 29, 2000, were used to conduct this study.

4.3. Operationalization of Key Concepts

To assess the hypotheses regarding the impact of cognitive processing limits on Usenet virtual public discourse, it is necessary to operationalize two key concepts, *message complexity* and *interactive group*, which in turn requires that the chain of interactions between users be determined and measured. Further, the notion of an *interactive group* also implies a specified time frame within which the interactions between group members occur.

By *message complexity* we refer to the effort required by users to create and/or read a message. Common sense informs us that on average the effort required to create a Usenet message will correlate with a number of message characteristics, the most obvious being message length. For each Usenet message the following variables related to length were calculated: the number of words in the body of the message (words); the number of words on nonindented lines in the body of the message (new words); the number of lines according to the header field, which typically includes lines of attachments; the number of lines excluding those of attachments (lines); and the number of nonindented lines excluding those of attachments (new lines).

It is assumed that, to some degree, these variables correlate with message complexity (shorter messages on average being simpler). Of course, none of these measures are ideal because the effort required to write a message relates to many other factors. Examples include the concepts the author is intending to convey, the context of the message in the discourse stream, and the complexity of the language required. A more refined measure could be computed via the use of a more complex algorithm. Variables utilized in such an algorithm could include sentence length, technical word use, and recognition of the extent to which a message was impacted by being part of a discussion thread. However, we believe that for our purposes such analysis is not called for.

The *size of interactive discussion groups* can be conceived of in a number of ways, so it is important that we examine how the term should be operationalized. Is it the size of the group the number of subscribers, the number of contributors, the number of messages, the number of contributors of interactive and/or reactive messages, or some other measure? Obviously, the number of newsgroup postings and the number of newsgroup contributors will be highly correlated. However, it makes sense to distinguish posters of responses from purely one-way posters, because one-way posters may not actually be engaged in group discourse and because the act of responding to another individual's message appears to be qualitatively different. Further, group size is generally understood in terms of people, not their output. Based on this logic and for the sake of simplicity we will operationalize the *size of the interactive discussion group* in this paper as the number of contributors of reply messages (interactive and/or reactive messages). This requires that we identify reply messages and message threads. This cannot be achieved simply by using the message header references because these references are not particularly accurate (Lewis and Knowles 1997, Smith 1999). As a result, the highly difficult task of thread reconstruction of the millions of the messages collected had to be undertaken. We describe how we achieved this objective in detail in the appendix.

What is the time period used to operationalize the concept *interactive group*? Should this be a day, week, month, or some other duration? At the time of data collection Usenet servers typically stored postings for about a week, suggesting that looking at a weekly set of data would more closely approximate user interactions with newsgroups. Hence, an examination by week has some face validity because it probably relates closely to how typical users perceive newsgroups. Further, data plots suggest that for some of the issues examined in this paper, looking at the data from a monthly perspective results in low granularity. On the other hand, daily and weekly plots look quite similar, despite some quite strong daily fluctuations. All these points suggest that weekly measures are most appropriate for this study. Therefore, for the current purposes, interactive discussion group size will be considered equivalent to the number of interactive or reactive posters to a newsgroup over a one-week period.

Active newsgroup participation or *newsgroup stability*: An examination of the Usenet data shows that many active newsgroup users do not post every week, suggesting that the measure of stability or active participation should be based on a longer time period. It therefore seems reasonable to examine this issue by choosing a month-to-month scale to measure stability. Therefore, for the purposes of this study, we define *proportional poster stability* as the percent of posters in a month that also posted in the previous study month.

A final term that needs to be operationalized is the concept of *operating at AvMaxCL*. Because this is the key theoretical construct our research is attempting to demonstrate, it cannot be directly operationalized. Instead, it is assumed here that a large proportion of Usenet newsgroups operate at this level of communication load, and it is hypothesized that we will be able to observe its impact by large-scale observation of Usenet interactions.

4.4. Data Analysis

The statistical analysis was conducted using SPSS. For all results reported in this study, the significance level deemed appropriate was the smallest value reported by the SPSS package for the analysis in question. For large-scale analysis this was $p < 0.0000$, and for analysis with hundreds of variables this was $p < 0.000$. Regression analyses were conducted to test each hypothesis. For each regression analysis backward elimination was used. All the variables included in the final models reported are listed in the results section. Further, examinations were made for multicolinearity, and probability plots were produced to examine if outliers should be removed from derived models.

To test the first hypothesis that message length will relate to group size, we need to deal with the problem of *regression toward the mean*, where larger groups have less variance of mean measures (Newell and Simpson 1990). This effect was expected because word and line count cannot be negative, so outliers are going to have a larger impact on the means of small groups, resulting in a decreasing slope as group size increases. This means that approaches such as orthogonal polynomials would simply demonstrate the existence of a curve that could be explained as a statistical artifact. The method we chose to address this problem was to compare the transformed ranks of the sizes of individual messages rather than aggregations. This was achieved by ranking the entire sample of messages according to each of the measures of comparative message complexity and then transformed to approximate normality using the Rankit proportion estimation formula ($(r-1/2)/w$) (where w is the number of observations, and r is the rank, ranging from 1 to w, with tied ranks allowed) (Chambers et al. 1983). Using these new variables, it is possible to conduct regression modeling to assess if message length was related to group size, while taking into account concerns about *regression to the mean*.

To test the second hypothesis that shorter messages were more likely to seed discourse, we used a standard logistic regression, with its underlying nonlinear mathematical model. To test the final hypothesis regarding newsgroup stability, it was necessary to remove outliers to derive a model fully meeting requirements for normality.

4.5. Data Preparation

As discussed in §4.3, to operationalize the concept *size of interactive discussion groups*, and to test hypotheses related to our theoretical model, we need to identify messages' positions in discussion threads. In the appendix, we detail the techniques used to achieve this. It should be noted that the aim of this data preparation was not to perfectly identify reply messages and threads in this multimillion message data set, but rather to identify a sufficient percentage of reply messages and message threads so as to enable hypothesis testing with face validity. We considered this to be the case if the measures of specificity (the probability of a negative result for a true negative) and sensitivity (the probability of a positive result for a true positive) were above 90%, because we believed this to be enough to enable observation of the hypothesized effects of information overload in a large data set. In the end we were able to identify *reply messages* with a specificity of 96% and a sensitivity of over 99% and *parent messages of a message-reply pairs* with a specificity of 92.4% and sensitivity of 99%.

5. Results

With the operationalization of key terms and the demonstration of both the adequate identification of replies and the reconstruction of threads, it is possible to examine the hypothesized effects of the information overload on virtual public message dynamics.

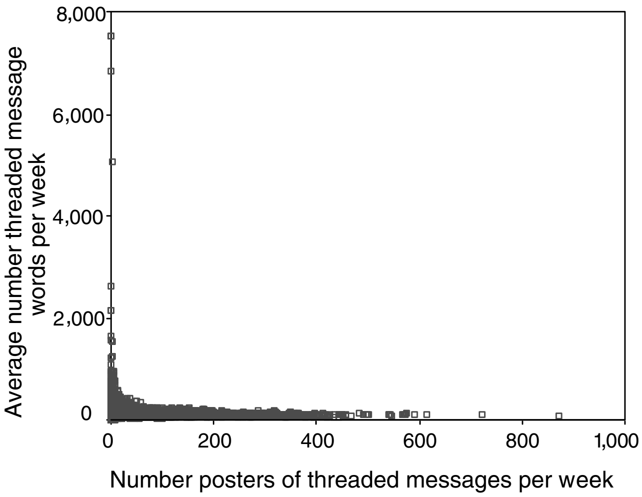
Hypothesis 1. *At AvMaxCL, as the number of communications increases, there will be a decrease in the average complexity of response messages until asymptote.*

It is assumed here that a sufficiently large proportion of Usenet newsgroups operate at AvMaxCL, and therefore we will be able to observe its impact by large-scale observation of Usenet interactions. Thus, this hypothesis is explored by examining if average complexity of response messages decreases as the size of the interactive Usenet newsgroup discussion group increases. Scatter plots of the various message complexity measures against group size/activity measures were made to provide insight into the appropriateness of various statistical methods and assure face validity to the hypotheses under examination. One of these plots is presented in Figure 3.

Figure 3 is a scatter plot of the average number of words in threaded messages (replies) by the number of posters of such messages. Each of the various scatter plots of message complexity by the size of the interactive discussion groups look similar. They also display the expected relationship between the size of the interactive newsgroup and various measures of message complexity.

Table 1 shows the comparison between first and fourth quartiles of group size (number of posters of threaded messages) for various measures of message complexity. For all measures, the average message complexity is reduced as the group size increases, and this is the case for all quartiles examined and for both measures of group size.

Figure 3 Plot of Average Number of Words in Threaded Messages Posted to Newsgroups by the Number of Posters of Threaded Messages



Unfortunately, while the plots and quartile data support the model proposed, they do not provide strong evidence. This is because, as noted in the methods section, in statistical terms the plots and the tables can be explained in terms of *regression toward the mean*, where larger groups have less variance of mean measures (Newell and Simpson 1990). Because word and line count cannot be negative, outliers are going to have a larger impact on the means of small groups, resulting in a decreasing slope as group size increases. To side-step the problem of regression to the mean, we chose to look for the hypothesized effects by examining the average of ranks of individual messages rather than aggregations. This was achieved by ranking the entire sample of threaded messages according to each of the measures of comparative message complexity and then transformed to approximate normality using the Rankit proportion estimation formula (see §4.4). The rankings were from 1 to 1,499,133. Variables were then computed and matched to individual messages regarding the newsgroup activity during the study week messages were posted. Using these new variables, it is possible to use regression to see if the number of posters, and/or number of interactive threads, relates to *message complexity* without the concern of regression to the mean. This approach also allowed us to account for factors such as major newsgroup topical classification ("Comp.," "Misc.," etc.) and message crossposting.

Table 1 Message Complexity Measures for First and Fourth Quartile of Distribution of Posters of Threaded Messages

Variables	Quartiles	Number	Mean	Std. deviation
Threaded message lines per week computed from message header data	first quartile	3,164	36.14	62.46
	fourth quartile	3,072	30.51	8.24
Threaded message lines per week computed from content	first quartile	3,164	34.39	61.89
	fourth quartile	3,072	28.88	8.04
Number of nonindented threaded message lines per week	first quartile	3,164	20.49	52.03
	fourth quartile	3,072	16.49	3.93
Number of threaded message words per week	first quartile	3,164	208.68	353.42
	fourth quartile	3,072	182.96	55.01
Number of words on nonindented threaded message lines per week	first quartile	3,164	126.25	275.83
	fourth quartile	3,072	106.70	29.03

Unfortunately, while the ranking and variable matching enabled regression modeling, this approach results in a loss of variance and predictive/explanatory power. In other words, the ranking approach with transformation enabled a regression to be run, although this approach meant that the model's ability to explain variance was not particularly high. This was considered acceptable. It could be used to assess the notion that group activity related to message complexity, although it does not provide insight into the strength of any observed relationship between newsgroup activity and message size.

Each of the five possible measures of message complexity were used as the dependent variable in different regression models. Of the five measures, the one most strongly predicted by group size was the average number of message lines calculated by the posters' client newsreader ($F = 14{,}836.24$; df = 9, 1,499,124; $p < 0.0000$). The regression modeling suggested that the newsgroup size (number of threaded messages posted or number of threaded posters) did predict message length (shorter messages being posted to more active groups). The variables in the final model with Rankit codification of number of lines as the dependent variable were as follows.

• The number of messages posted to the newsgroup in the week the message was posted—The results show that as predicted the more active the newsgroup, the smaller on average the messages posted.

- A dummy variable for each major Usenet newsgroup categories ("Soc.," "Comp.," etc.) because the categories to some extent predict the content and length of the messages they contain.
- If the message was crossposted (sent simultaneously to multiple newsgroups) because these messages were longer on average. This can perhaps be explained by the fact that such messages need to contain enough content to be comprehensible to multiple audiences.
- Messages' position/depth in a discussion thread; messages that were deeper in a discussion thread had more lines on average, which is not surprising because they needed to reference earlier aspects of threaded discussions.

Hypothesis 2. *At AvMaxCL, simple group CMC messages will be more likely to generate responses than complex messages.*

It is assumed here that a large proportion of Usenet newsgroups on occasion operate at AvMaxCL and that, therefore, we will be able to observe its impact by large-scale observation of Usenet interactions. There were 593,019 messages that could be considered true unambiguous broadcasts or one-way messages in our sample of Usenet messages. From this sample of one-way messages, 255,697 were found to have initiated (seeded) discussion within their newsgroup during the study period. Table 2 below compares the means for various message complexity measures and average newsgroup crossposting between broadcast messages that seeded and those did not seed further discussion.

Table 2 Means Table of Broadcast Message Complexity and Discourse Seeding

Variables	Seeds/does not seed discourse	Number	Mean	Std. deviation
Threaded message lines per week computed from message header data	one-way	337,322	62.29	645.66
	seeds thread	255,697	24.50	225.34
Threaded message lines per week computed from content	one-way	337,322	53.53	154.55
	seeds thread	255,697	22.53	51.01
Nonindented threaded message lines per week	one-way	337,322	53.22	154.41
	seeds thread	255,697	22.34	50.75
Threaded message words per week	one-way	337,322	318.57	958.13
	seeds thread	255,697	144.85	340.54
Words on nonindented threaded message lines per week	one-way	337,322	316.69	956.97
	seeds thread	255,697	143.60	338.46

Table 2 shows that as predicted broadcast messages that seed discourse are shorter than those that fail to seed discourse (all differences being statistically significant). They are also less likely to have been crossposted. Of course the analysis in Table 2 does not take into account the possibility that these differences are simply due to the fact that shorter messages are sent to newsgroups where all messages are more likely to be replied to (a pattern also predicted by the hypothesized model). For example, the "Comp." newsgroups, which are about various computer issues, are very active and contain shorter messages than recreational ("Rec.") discussion groups.

Logistic regression on the binary outcome "seeded/did not seed" discussion thread was carried out to assess this possibility. The logistic model that best describes the dynamics of discourse seeding was able to predict 63.56% of the cases with a Wald χ^2 of 56,635.936; $\text{df} = 9$, $p < 0.0000$. The variables in the final model as predictors of new messages seeding a discussion thread were as follows.

- Number of message lines (all the measures of message complexity such as number of words were examined separately to avoid multicolinearity, and they were all similar); As predicted, this was found to negatively correlate with seeding discourse (more lines are less likely to seed).
- Overall activity of the newsgroup (measured by the number of messages posted per week); new messages posted to more active newsgroups were more likely start a discussion thread. This in part conforms to Palme's (1995) communication response function.
- If newsgroup was moderated; when newsgroups were moderated, new messages were less likely to seed a discussion thread than in the unmoderated situation where fewer constraints to inter-user interactivity exists.
- Crossposting—If it was crossposted, it was less likely to receive a reply in an individual newsgroup; this is not surprising because a crossposted message is not directed to any particular newsgroup.
- Dummy variables for each of the major newsgroup categories being examined ("Soc.," "Comp.,"

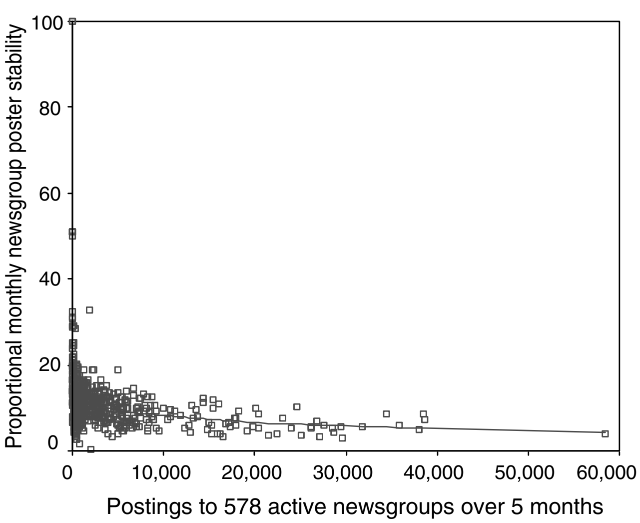
etc.) because the categories to some extent predict interactivity of newsgroup discussions.
The findings of the logistic regression modeling argue strongly for the conclusion that smaller messages are more likely to generate ongoing discourse.

Hypothesis 3. *At AvMaxCL, as group CMC complexity increases, there will be an increased tendency for individuals to end or reduce active participation.*

It is assumed here that a large proportion of Usenet newsgroups operate at AvMaxCL and that, therefore, we will be able to observe its impact by large-scale observation of Usenet interactions. As discussed in the methods section, we define proportional poster stability as the percent of posters in a month that also posted in the previous study month. This allowed for the examination of user stability over a five-month period.

Figure 4 displays the number of messages posted to the 578 newsgroups that were active during the first 5 months of the study. Only 11.5% of posters sent messages 2 months in a row. Because of the constraints imposed by the proportionality of the stability measure (0 to 100) it seems reasonable in this case to also plot on Figure 4 a regression line to highlight the reduction in stability as newsgroup activity increases. The drop in the proportion of individuals involved in sustained discourse is quite strong, with a Spearman's rank correlation coefficient of -0.43 ($p < 0.000$; $n = 565$).

Figure 4 Proportional Newsgroup Poster Stability



To remove the outliers seen on the plot, which appear to result from the small user populations of some of these groups, rank correlations were also conducted using the top third of the sample (those newsgroups months with more than 2,957 messages posted to them, accounting for 1,943,343 of the studies messages). Using this smaller sample the Spearman's rank correlation coefficient was -0.47 ($p < 0.000$; $n = 192$), highlighting the strength of this finding.

Regression modeling was conducted with newsgroup stability as the dependent variable. The following variables were included in the final models.

- Number of posters; as predicted, the more active newsgroups were more unstable.
- Newsgroup moderation; moderated newsgroups were more stable, but this is not surprising because they were smaller and probably have a more dedicated core of users.
- Crossposting; newsgroups with crossposted messages had less stable membership. This again suggests, like moderation, that focused discussion results in more stable newsgroup membership and activities.
- Dummy variables for each of the major newsgroup categories being examined ("Soc.," "Comp.," etc.) because the categories to some extent predict interactivity of newsgroup discussions.

If outliers are included in regression modeling and the moderate skew of the proportional stability measure is ignored so that the full sample can be modeled, then the number of posters is found to be the best predictor of newsgroup stability ($R^2 = 0.24$; $F = 22.1$; df $= 8{,}556$; $p < 0.000$). Full sample regression modeling in this situation is not ideal because the outliers result in a nonnormal distribution of regression residuals. It was therefore decided to run a regression on the top third of the sample because this removes the outliers that probably result from small group size, and results in a normal distribution of regression residuals. The outcome for this final model was similar to that, although the discriminatory power of the regression was greatly improved ($R^2 = 0.44$; $F = 17.7$; df $= 8{,}183$; $p < 0.000$), with group size being the best predictor of proportional poster stability.

6. Discussion and Conclusions

A large proportion of scholarly examination of the public online behavior of Internet users has been in terms of *virtual community*, using social theory and generally adopting ethnography-like or social network approaches. These approaches underplay the relationship between online interaction spaces and the user behavior they host. Information system research needs to examine the nature of the relationship between the virtual spaces typically used for public online behavior, their technological platforms, and the behaviors such systems contain.

A theoretical framework and methodology for researching online interpersonal interaction spaces and the technologies that support them has been proposed. The theory indicates a dynamic interplay between volume and complexity of information that is conducted in online groups and the participation and responses of users in these spaces. The methodology used to test this theory utilizes large-scale field studies into user behavior in online spaces to identify technology-associated constraints to sustainable patterns of virtual public interaction dynamics. Empirical research was undertaken to assess the validity of both the theoretical framework and methodology.

Overall, the empirical findings support the assertion that individual information-overload coping strategies have an observable impact on mass interaction discourse dynamics. Evidence was found for the three hypotheses examined in this regard: (1) users are more likely to respond to simpler messages in overloaded mass interaction; (2) users are more likely to end active participation as the overloading of mass interaction increases; and (3) users are more likely to generate simpler responses as the overloading of mass interaction increases.

While we do not believe that our findings rule out various other possible explanations, they do highlight the utility of the method and theory. The full research program (Lakatos 1978) as outlined by Jones (2002) makes novel predictions that can be examined empirically. This is demonstrated by this paper and by Liu (1999) and Ekeblad (1999). Further, the methodology outlined can be used for comparing various virtual public technologies (see Jones 2003 for a comparison between Usenet newsgroups and e-mail lists). So, while we call for further validation of the explanatory model, the significance of the results and method stand in their own right.

The emergence of mass interaction has presented new opportunities to learn about system effects on mediated group discourse. Empirical research into the *systemic* nature of the patterning of social relationships in cyberspace has, despite its importance, been relatively rare. Prior research based on a systems approach did examine Internet group communication, such as critical mass limitations (Markus 1987, Rafaeli and LaRose 1993), the modeling of free riding using the Napster-like Gnutella network (Adar and Huberman 2000), disintermediation (Rafaeli and Ravid 2003), modeling the interrelationship between homepages (Adamic and Adar 2000), exploring the self-organizing nature of e-mail lists (Ekeblad 1999), and analyses of the structure of the World Wide Web and Internet flows (Adamic 1999, Barnett et al. 2001, Park et al. 2002). The work described here is one of the first attempts to empirically explore the impact of systems effects in Usenet discourse.

We have argued in this paper that the recognition of the systemic nature of virtual public discourse allows for the examination of CMC technologies in terms of group-level usability. Currently, it is widely accepted that "reliable measures of overall usability can only be obtained by assessing the effectiveness, efficiency and satisfaction with which representative users carry out representative tasks in representative environments" (Bevan and Macleod 1994, p. 132). The current, dominant view supports a focus on individual level analysis, usability laboratories, and ethnographic methods for context. Without discounting the value of these approaches, the methodology presented in this paper represents an alternative, unobtrusive aggregate approach.

We see this research progressing along several trajectories. First, we expect to see more work on large-scale comparative analysis of virtual public discourse dynamics for a variety of technologies (Listserv, IRC, Web-based bulletin board systems, etc.). Second, there is potential for a thorough empirical examination of the impact of the various ways individuals can respond to overloaded virtual public discourse. Third, further research into alternative explanations for the underlying causes of consistent patterns of virtual public mass interaction would be welcome. Fourth, research could progress through various methodological and theoretical refinements that would result from an examination of related issues such as time scaling and longitudinal impacts (Schoberth et al. 2003). Finally, through the utilization of the knowledge gained into group-level usability, it should be possible to design better CMC tools.

The combination of theory and methodology outlined here and the associated empirical research contribute to the understanding of the dynamics of online group interaction. The theory made verifiable predictions as to the nature of online behavior; specifically, that the impact of individual cognitive processing limits can be observed in the mass interaction dynamics of virtual publics. Overall our theoretical model, methodology, and empirical results suggest a new way of exploring how CMC technologies affect the discourse they host.

Appendix. Identifying Replies and Parent Messages

Identifying Replies

A range of tools and techniques can be used to identify messages as replies, including data mining and machine learning. Here we constructed an algorithm utilizing the degree to which reply indicators in the subject line, the message body, and the message headers each correctly identified a message as a reply. We deemed the algorithm used adequate for the task at hand because it was above our 90% benchmark for sensitivity and specificity.

Subject Line Indicators: The number of standard reply indicators (“reply,” “(re),” or “re:”) that appeared in each message’s subject line were counted. Of the 2,652,552 Usenet messages examined, 2,042,290 messages (77%), contained “re:” or “(re)”. Only 2,238 messages subject lines contained the word reply. Out of 2,042,290 messages with any of these reply indicators in the subject line, 100 messages were chosen at random for examination by two reviewers to determine if they were replies. These reviewers were simply asked to categorize each of these messages as either replies or new messages. Both reviewers concluded that 100% of these messages were replies. There were only a small percentage of messages with indications that they were forwarded. There were 1,059 subject lines that contained “FW,” 1,854 that contained “FWB,” 1,370 that contained “FWD,” and 4,240 that contained the original message. These 8,523 messages represented 0.3% of the messages. There were 100 messages with forwarding indicators that were chosen at random for examination by two reviewers to determine if they were also reply messages. Both reviewers concluded that 67% of the forward messages were clearly replies, and for 3% the status was unclear (inter-rater reliability $\alpha = 1$).

Message Body Indicators: Messages often refer to early messages by containing a line with the format similar to: “Name ⟨e-mail address of earlier poster⟩ wrote:”. Therefore the number of times that messages contained text with the text strings “wrote:”, “write:”, “wrote;”, or “write;” was counted. In total 1,201,541 messages (45%) contained such strings. There were 100 messages with message body text, indicating they were replies chosen at random for examination by two reviewers. Both reviewers concluded that 99 of these 100 messages were replies (again with inter-rater reliability $\alpha = 1$). The most common indenting used in Usenet messages to signify text quoted from an earlier message is “>”. Multiple indenting indicates that it is a quote from a message that was itself a quote, so depth of indentation is an indication of thread depth. For each message, the number of lines that began with indenting “>” were examined. The number of these lines that were followed by one to four further indentation marks was also examined (number of lines that started with “⋙” or “⋙”). We found that 1,514,297 (57%) of messages had more than two lines with “>” indentation. A message that contained more than two lines of indenting was considered a reply because an examination of many messages with one or two lines of indentation showed a large percentage to be unrelated to quoting an early message. There were 100 messages with an indenting measure that suggested they were replies chosen at random, and they were then examined by two reviewers. The reviewers concluded that 99 of these 100 messages were replies (inter-rater reliability $\alpha = 1$).

Message Header Indicators: Usenet headers have a field called “references:” or more rarely “in reply to:” where the “message-ID’s” of the messages to which they are “replies” are stored. A total of 1,749,532 messages (66% of study sample) contained references to other messages. Eighty three percent of the message-ID’s contained in these header fields referred to messages posted to the same discussion group within the study period. The average time between message pairs linked together by reference headers was 1 day; 90% of message-pairs were separated by less than 2.5 and 99% of responses occurred with the first two weeks. There were 30,243 messages containing an “in reply to:” field with message-ID’s (1% of the messages sampled). Out of 1,749,532 messages with references, 100 messages were chosen at random for examination by two reviewers to determine if they were replies. Both reviewers concluded that 100% of these messages were replies (inter-rater reliability $\alpha = 1$).

Reply Message Identification Algorithm: The algorithm used to identify reply messages took into account all the reply indicators discussed above, which were all highly correlated (84% of messages with “re:” or “(re)” in the subject line also have message references). Our plan was to iteratively improve our initial algorithm until it either reached an acceptable level of sensitivity and specificity, or it was concluded that an alternative approach was needed. For our initial analysis the algorithm we used was computed as follows: If the subject line contained “re:” or “reply:”, it was considered a reply because this represented the largest group of messages and was found to be strongly indicative of a message being a reply. If a reply string was not found in the subject, then messages were required to have a score of 0.8 or higher to be considered as a reply. More than two lines with indenting result in a weighting of 0.6. "In reply to:" was given a weight of 0.5. Having message content that indicated that it was forwarded was given a weight of 0.4. Finally, a message reference was given a weight of 0.3. This formula takes into account the large percentage of messages with subject lines indicating that they are replies, and the fact that the vast majority of messages with referencing also have reply strings in the subject. The exact weights were chosen heuristically. In the end 2,061,179 messages (78% of the study sample) were coded as replies.

Initially, 100 messages coded as replies were chosen at random for examination by two reviewers to determine if they were indeed replies. These reviewers were simply asked to categorize each of these messages as either replies or new messages. Both reviewers concluded that 100% of these messages were replies.

A second sample of 500 messages coded as replies were chosen at random for examination by two additional reviewers, and again they found that 100% of these messages were replies. There were 100 messages that were coded as not being replies and were chosen at random for examination by two reviews to determine if they were indeed not replies. Four of these messages were falsely coded. For three of the messages the status was unclear because even after reading the messages in context, it was not possible to tell whether they were replies. Again, a second sample of 500 messages coded as nonreplies were chosen at random for examination by two additional reviewers. They both found 19 of the messages to be incorrectly coded; 1 was unable to code 14 and the other 15 of the messages. Using these results from the human reviewers, it was concluded that the measure has a specificity of 96% (the probability of a negative result for a true negative). The weighted measure's specificity was higher than it would be had we used an unweighted approach, so it was deemed adequate for the task.

Using Bayes' Theorem $(0.999 * 0.7888)/((0.999 * 0.788) + (0.04 * 0.2112))$, it can be concluded from the above that the chances of a reply being correctly labeled was 99%. This is based on the assumption that if a larger sample of messages was used, the true positives would have dropped below 100% to around 99%. This number is fairly robust because the large majority of messages are replies.

Identifying Parent Messages

Message header referencing is not a realiable means of indentifying and reconstructing discussion threads (Lewis and Knowles 1997). Further, 16% of messages coded as replies did not have references, which means that threading could not be reconstructed by solely relying on the references contained in the message headers. To identify parent messages we proceeded as follows: First, for each reply message with a "reference:" header, a search was made of the postings to its newsgroup during the study period to determine if a message-ID referenced as the parent message was posted to the same newsgroup during the study period. If a match was found, then the parent-child message relationship was noted. Second, if the parent was not found, then a similar step was taken for the "in reply to:" field, and again the parent-child message relationships found were noted. Third, various searches were made to match the subject line after reply indicators such as "re:" had been removed. These subject string searches first checked backward sequentially for 14 days and then forward half a day. Further, around 25% of reference-based replies were to replies that, according to the client newsreader used, were actually posted after the message replies when these times were adjusted to GMT (most of the incoherent time relationships were a matter of minutes, which can be explained by inaccurate clocks on client computers). The parent-child message relationships that were found using the subject line "re:" were noted. Finally, the Oracle SQL extension "connect by" command was used to construct the discussion threads for all the newsgroups from the parent-child (message-reply) pairs. This Oracle command is used to construct and select data based on hierarchical relationships (it builds tree relationships). Using this command, it is possible to reconstruct the discussion threads using the identified pairs of relationships and to ensure that, for example, a child is not also its parent's parent.

Of the 2,061,179 reply messages in the study sample, it was possible (using the above technique) to identify a parent message from the same newsgroup study sample for 1,502,991 messages (73% of reply messages). Interestingly, for those messages with references, it was possible to identify their parent message in 83% of the cases. Of the 1,502,991 message-reply pairs contained in the database, 100 were extracted at random. Two reviewers then examined each pair to see if the parent message did indeed look like a parent message. Of these messages, 100% were considered correctly paired by the human raters. This process was repeated by 2 additional reviewers with an additional 100 message-reply pairs with the same outcome. So the specificity of parent message identification appears to be over 99%. When the Oracle SQL extension "connect by" command was used to construct the discussion threads for all the newsgroups from the 1,502,991 message-reply pairs, 3,857 messages (0.26%) were found to be in conflict with other message pairs.

Four reviewers examined all the messages to one newsgroup chosen at random to assess the adequacy of the parent search algorithm. The newsgroup selected received 1,367 postings during the study period; of these there were 170 reply messages for which no parent messages were found. Using this time-consuming manual search process, our human reviewers were able to manually find a parent message for approximately 39% (67 out of 170) of the orphaned messages. This was achieved by reviewers manually searching for each orphaned message's parent (inter-rater reliability $\alpha = 0.9878$—The only situation in this study in which the outcome was not an initial 100% agreement between raters). At the end they compared their findings with other reviewers to come to a consensus about the 2–3 messages for which there was an inconsistency.

These findings suggest that the true percentage of replies with parent messages in the sample is around 83%, and that approximately 13% of the parent messages were not identified. In other words, approximately 87% of parent messages were identified. If we consider the probability of falsely having a positive test result as equivalent to stating that a parent message does not exist when it actually does (0.4), then the specificity of the algorithm is 92.4%. We recognized that the true percentage may vary because only one newsgroup was examined using labor-intensive manual verification. However, as the aim was simply to demonstrate that threads were identified adequately for hypothesis testing, we were satisfied with this level of data preparation, considering our finding of over 99% sensitivity in parent message identification.

References

Ackerman, M., L. Palen. 1996. The Zephyr help instance: Promoting ongoing activity in a CSCW system. *Proc. ACM Conf. Human Factors Comput. Systems*.

Adamic, L. A. 1999. The small world web. *Proc. 3rd Eur. Conf. Digital Libraries*, 443–452.

Adamic, L. A., E. Adar. 2000. Friends and neighbors on the Web. PARC Xerox Manuscript 1501. Palo Alto, CA, http://www.parc.xerox.com/istl/groups/iea/papers/web10/index.html.

Adar, E., B. Huberman. 2000. Free riding on Gnutella. *First Monday* **10**(5).

Barnett, G. A., B. Chon, H. W. Park, D. Rosen. 2001. Network analysis of international Internet flows. *Internat. Sunbelt Soc. Network Conf.*, Budapest, Hungary.

Bevan, N., M. Macleod. 1994. Usability assessment and measurement. M. Kelly, ed. *The Management and Measurement of Software Quality.* Ashgate Technical/Gower Press,

Bradner, E., W. Kelloggand, T. Erickson. 1999. The adoption and use of babble: A field study of chat in the workplace. *Proc. ECSCW.* Copenhagen, Denmark.

Broadbent, D. 1958. *Perception and Communication.* Pergamon Press, London, U.K.

Butler, B. 2001. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Inform. Systems Res.* **13**(4).

Chambers J. M., W. S. Cleveland, B. Kleiner, P. A. Tukey. 1983. *Graphic Methods for Data Analysis*. Wadsworth International Group, Duxbury Press, Boston, MA.

Cherny, L. 1999. Conversation and community: Chat in a virtual world. Center for the Study of Language and Information, Stanford,

Ekeblad, E. 1999. The emergence and decay of multilogue: Self regulation of a scholarly mailing list. *Proc. Eur. Assoc. Res. Learning Instruction.* Sweden.

Fletcher, R. J. 1995. *The Limits of Settlement Growth: A Theoretical Outline*. Cambridge University Press,

Forrester, J. W. 1969. *Urban Dynamics*. MIT Press, Cambridge, MA.

Fulk, J., J. A. Schmitzand, D. Schwarz. 1992. The dynamics of context-behavior interactions in computer-mediated communication. M. Lea, ed. *Contexts of Computer-Mediated Communication*. Harvester Wheatsheaf, New York, 7–29.

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

Herring, S. 1999. Interactional coherence in CMC. *Proc. 32nd Hawaii Internat. Conf. System Sci.*, Hawaii.

Hiltz, S. R., M. Turoff. 1978. *The Network Nation: Human Communication via Computer*. Addison-Wesley Publishing Company, Inc., London, U.K.

Hiltz, S. R., M. Turoff. 1985. Structuring computer-mediated communication systems to avoid information overload. *Comm. ACM* **28**(7).

Jones, S. 1995. Cybersociety: Computer-mediated communication and community. *Understanding Community in the Information Age*. Sage, Thousand Oaks, CA, 10–35.

Jones, Q. 1997. Virtual-communities, virtual-settlements and cyber-archaeology: A theoretical outline. *J. Comput. Mediated Comm.* **3**(3).

Jones, Q. 2002. The boundaries of virtual communities. Unpublished Ph.D. thesis, University of Haifa, Israel.

Jones, Q. 2003. *Appling Cyber-Archaeology, ECSCW 2003*. Kluwer Academic Publishers. In Press.

Jones, Q., S. Rafaeli. 2000a. Time to split, virtually: "Discourse architecture" and "community building" as means to creating vibrant virtual publics. *Electronic Markets: The Internat. J. Electronic Commerce Bus. Media* **10**(4) 214–223.

Jones, Q., S. Rafaeli. 2002. What do virtual 'tells' tell? Placing cybersociety research into a hierarchy of social explanation. *Proc. 33rd Hawaii Internat. Conf. System Sci.*, IEEE Press.

Kling, R., E. Gerson. 1977. The social dynamics of technological innovation in the computing world. *Symbolic Interaction* **1**(11) 132–146.

Lakatos, L. 1978. *The Methodology of Scientific Research Programmes: Philosophical Papers*. Cambridge University Press, Cambridge, U.K.

Lea, M., R. Spears. 1991. Computer-mediated communication, de-individuation and group decision making. *Internat. J. Man-Machine Stud.* **34** 283–301.

Lewis, D. D., K. A. Knowles. 1997. Threading electronic mail: A preliminary study. *Inform. Processing Management* **33**(2) 209–217.

Licklider, J., R. Taylor. 1968. The computer as a communication device. *Sci. Tech.*

Liu, G. Z. 1999. Virtual community presence in Internet relay chatting. *J. Comput. Mediated Comm.* **5**(1).

Markus, M. L. 1987. Towards a "critical mass" theory of interactive media: Universal access, interdependence and diffusion. *Comm. Res.* **14** 491–511.

Miller, G. A. 1956. The magical number seven, plus or minus two: Some limits on our capacity to process information. *Psych. Rev.* **63** 81–97.

Newell, D., J Simpson. 1990. Regression to the mean. *The Medical J. Australia* **153**(August 6) 166–168.

Palme, J. 1995. The optimal group size in computer mediated communication. *Electronic Mail.* Artech House Publishers, London, U.K.

Park, H. W., M. Thelwall. 2003. Hyperlink analyses of the World Wide Web: A review. *J. Comput.-Mediated Comm.* **8**(4).

Rafaeli, S. 1988. Interactivity: From new media to communication. *Sage Annual Review of Communication Research: Advancing Communication Science*, Vol. 16. CA, 110–134.

Rafaeli, S., R. J. LaRose. 1993. Electronic bulletin boards and 'public goods' explanations of collaborative mass media. *Comm. Res.* **20**(2) 277–297.

Rafaeli, S., G. Ravid. 2003. Information sharing as enabler for the virtual team: An experimental approach to assessing the role of electronic mail in disintermediation. *Inform. Systems J.* **13** 191–206.

Rafaeli, S., G. Ravid, V. Soroka. 2004. De-lurking in virtual communities: A social communication network approach to measuring the effects of social and cultural capital. *Proc. 37th Hawaii Internat. Conf. System Sci.*, Big Island, Hawaii.

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

Rogers, E. M., R. Agarwala-Rogers. 1975. Organizational communication. G. L. Hanneman, W. J. McEwen, eds. *Communication Behaviour.* Addision Wesley, Reading, MA, 218–236.

Schmitz, J., J. Fulk. 1991. Organizational colleagues, media richness, and electronic mail: A test of the social influence model. *Comm. Res.* **18** 487–523.

Schoberth, T., J. Preece, A. Heinzl. 2003. Online communities longitudinal analysis of communication activities. *Proc. 36th Hawaii Internat. Conf. System Sci.*, Big Island, Hawaii.

Simpson, R., A. Renear, E. Mylonas, A. van Dam. 1996. 50 years after "as we may think": The Brown/MIT Vannevar Bush Symposium. *Interactions* **3**(2).

Smith, M. 1999. Invisible crowds in cyberspace: Mapping the social structure of the Usenet. M. Smith, P. Kollack, eds. *Communities in Cyberspace*. Routledge, London, U.K., 195–219.

Smith, M., A. Fiore. 2001. Visualization components for persistent conversations. *Proc. ACM's Conf. Human Factors Comput. Systems*, Seattle, WA, 136–143.

Spears, R., M. Lea. 1992. Social influence and the influence of the "social" in computer-mediated communication. M. Lea, ed. *Contexts of Computer-Mediated Communication.* Harvester Wheatsheaf, New York, 30–65.

Sproull, L., S. Faraj. 1997. Atheism, sex and databases: The Net as a social technology. S. Kiesler, ed. *Culture of the Internet.* Lawrence-Erlbaum Associates, Inc., Mahwah, NJ.

Sproull, L. S., S. Kiesler. 1986. Reducing social context cues: Electronic mail in organizational communication. *Management Sci.* **32**(11) 1492–1512.

Thorn, B. K., T. Connolly. 1987. Discretionary databases: A theory and some experimental findings. *Comm. Res.* **14**(5) 512–528.

Wellman, B., M Gulia. 1999. Virtual communities as communities: Net surfers don't ride alone. P. Kollock, M. Smith, eds. *Communities in Cyberspace.* Routledge, New York.

Whittaker, S. 1996. Talking to strangers: An evaluation of the factors affecting electronic collaboration. *Proc. CSCW '96*, Cambridge, MA, 409–418.

Whittaker, S., L. Terveen, W. Hill, L. Cherny. 1998. The dynamics of mass interaction. *Proc. ACM's Conf. Comput. Supported Cooperative Work*, Seattle, WA, 257–264.