An Empirical Exploration of Mass Interaction System Dynamics: Individual Information Overload and Usenet Discourse

Quentin Jones* & Gilad Ravid[Ψ], Sheizaf Rafaeli[Ψ]
* The Department of Information Systems, College of Computing Sciences
New Jersey Institute of Technology, Newark, New Jersey
[Ψ]*Graduate School of Business University of Haifa, Haifa Israel*
Email: qgjones@acm.org, sheizaf@earthlink.net

Abstract

The large-scale adoption of computer mediated communication technologies has resulted in what has been described as "mass interaction", shared discourse between hundreds, thousands or more individuals. A number of theoretical papers have made the argument that because of the existence of various technological and psychological constraints, the forms that mass interaction takes, can, partly be understood in terms of system dynamics. In particular, it has been suggested that user information overload results in non-linear feedback loops which impacts on discourse structure. This paper describes an empirical examination of three hypothesized effects of such loops by the analysis of 2.65 million USENET messages posted to 600 newsgroups over a 6-month period.

Statistical analysis of the data demonstrated the existence of the hypothesized effects and support the assertion that individual 'information overload' coping strategies have an observable impact on mass interaction discourse dynamics. This in turn suggests that the usability of computer mediated communication technologies can be examined in terms of group-level usability.

1. Introduction

The exponential growth in recent years of telecommunication technologies has resulted in a new era of interpersonal communication [20]. Computer mediated communication (CMC) tools have altered both one-to-one and one-to-many communication. The growth of discourse systems where the audience is a significant source of media content as well as its primary receiver has resulted in what has been described as "mass interaction" [23], shared discourse between hundreds, thousands or more individuals. The term virtual community is commonly associated with such large-scale discourse. However, no single dominant definition of the term exists [14]. In fact, a number of authors dispute the existence of virtual communities [13]. Despite the significance of the phenomena commonly labeled *virtual community*, this term is problematic. To avoid this potential confusion the term *virtual public* is used in this paper [15]. Virtual publics are symbolically delineated computer mediated spaces, whose existence is relatively transparent and open, that allow groups of individuals to attend and contribute to a similar set of computer-mediated interpersonal interactions. Mass interaction generally takes place in virtual publics.

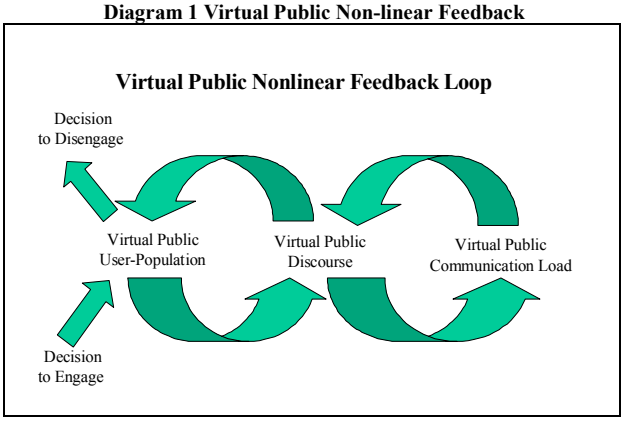
A number of theoretical papers have made the argument that because of the existence of various technological and psychological constraints, the forms that virtual public mass interaction takes, can, in part be understood in terms of system dynamics [7, 13-15, see 21 for a review]. More precisely it is argued that it is possible to consider discourse via virtual publics to be the output of a complex social system, using the notion of indeterminate hierarchies of explanation to coordinate the dependencies between levels [14,16]. Diagram 1 illustrates how the constraints acting on virtual public discourse result in non-linear feedback-loops. It works as follows: An increase in the membership of a virtual public will probably result in an increase in virtual public communication and user's average communication load. Communication load being the processing effort required by users to deal with a set of communications. However, it will not be possible for individuals to expand their involvement in virtual public communication indefinitely because of limits to the resources available to them to process group communication. Once virtual public communication becomes unmanageable or incoherent to individuals, then, the pattern of their involvement will alter, which in turn will impact on subsequent discourse dynamics.

Diagram 1 below highlights the argument that the cognitive processing abilities of groups are not simply the sum of its individual's cognitive processing capacities. It does this by describing active discourse engagement decisions in terms of the systems effects of communication load. It suggests that certain patterns of interactive mass-interaction cannot be sustained if the required processing effort is higher than the maximum amount individuals can or are prepared to invest. Therefore, it should be possible to observe empirically the

0-7695-1435-9/02 $17.00 (c) 2002 IEEE

combined effects of the average maximum communication load (AvMaxCL) individuals are prepared to process. In other words, mass interaction provides a unique opportunity to explore the impact of communication load on group discourse. This is because the large numbers involved allow broad statistical processes to be observed that may otherwise be hidden by differences between individuals and the social contexts of communication.

Diagram 1 Virtual Public Non-linear Feedback



Individuals can adopt a range of actions, or compensatory strategies, to reduce the impact of information overload resulting from group computer-mediated communication (CMC) [11]. These actions include:

1. Making an increased effort [5];
2. Learning new information management techniques [10, 11];
3. Failing to respond or attend to certain messages (e.g. focus on narrow topic);
4. Producing simpler responses;
5. Storing inputs and responding to them as time permits;
6. Ending active participation in the group communication [8]; and
7. Making erroneous responses.

All of these actions have a potentially observable effect on virtual public mass interaction, although changes in novel users' expertise, and user-effort, are likely to produce only short-term effects, which may be harder to identify indirectly via analysis of virtual public discourse dynamics. On the other hand, the other responses, if taken by a significant number of individuals, will impact on virtual public discourse in a more sustained and observable manner. We know that communication-processing load relates to a number of message-system characteristics. Users generally have to make more of an effort to reply coherently to a thread [17] than to a single message. Therefore, higher interactivity correlates with higher communication-processing load. Interactive communication refers here to the extent to which messages in a sequence relate to each other, and especially the extent to which later messages recount the relatedness of earlier messages [19]. Similarly, a high frequency of postings will require more processing by group members. Therefore, message frequency will also covary with communication-processing load. The relationship between discourse features such as number of interactive messages posted and communication load makes it possible to assess the basic principles of systems model. In other words, an increase in communication load at AvMaxCL will result in observable and therefore testable changes to system dynamics as group size grows. This is because, as the number of virtual public interactive messages / interactive posters increases in a situation where discourse is already overloaded, the systems model outlined above predicts that one would expect increase in the adoption by users of various compensatory strategies.

This paper assesses the implications of the above set of assertions through an empirical examination of three the hypothesized effects of overloaded mass interaction on discourse dynamics. These are that until asymptote, users are more likely to: 1) generate simpler responses as the overloading of mass-interaction increases; 2) respond to simpler messages in overloaded mass interaction; and 3) end active participation as the overloading of mass-interaction increases. These hypotheses are examined by analysis of Usenet discourse.

2. Research Methodology

2.1 Data Collection and Sampling

The Usenet is a system of electronic bulletin boards, referred to as newsgroups. It is not a computer network, but rather a network of bilateral agreements among system administrators to cooperate on bulletin board management [22]. Representative sampling of Usenet discourse is difficult; Whittaker et al's [23] 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 language text based conversational messages. For this project, data was collected from the 500 newsgroups studied by Whittaker et al enabling detailed historical comparisons. An additional 100 newsgroups were selected using Whittaker et al's approach with only minor modifications. This allowed for 100 moderated-groups to

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 IEEE

be selected. The full content of 3,293,995 postings were collected over eight months and stored in an Oracle database. The 2,652,552 messages collected over the 6-months from 1st August 1999 to 29th February 2000, were used to conduct this study.

2.2 Data Analysis

It is not possible to assess the impact of cognitive processing limits on virtual public discourse without first trapping interactions between users. Therefore a logical starting point for the analysis of the Usenet data is the extent to which message types (e.g. one-way or reply messages) were accurately identified and discussion threads successfully reconstructed. Such an assessment requires the measurement of two conditional probabilities in a similar fashion to the assesment of medical diagnostic tests [4]. These are the probability associated with a positive result for a true positive (sensitivity or recall), and the probability associated with a negative result for a true negative (specificity). Sensitivity (referred to as 'recall' in the information retrieval literature) and specificity are computed by Bayes' Theorem [12][1]. Using the medical analogy it is equal to:

$$\frac{P(T^+/D) * P(D)}{P(T^+/D) * P(D) + P(T^+/\not D) * P(\not D)}$$

Where: P(D) is the probability of truly having the disease.
$P(\not D)$ is the probability of not having the disease.
$P(T^+/D)$ is the probability of having a positive test result and having the disease.
$P(T^+/\not D)$ is the probability of falsely having a positive test result.

Once adequate sensitivity and specificity is demonstrated, the three hypothesized effects of the information overload non-linear feedback loops under examination can be assessed.

3. Results

3.1 Analysis of Thread Trapping

As noted above, trapping interactivity is essential. To achieve this end discussion threads need to be reconstructed reasonably accurately. Therefore, it is important to identify if a message is truly a "reply", and if it is, to correctly identify its "parent" message.

3.2 Identifying Replies

The algorithm used to identify reply messages took into account a variety of header fields, extent of message-body indentation, and reply and forward indicators in message body text.

Each message's subject line the number of times "reply", "(re)" or "Re:" appeared was counted. Of the 2,652,552 Usenet messages examined, an extremely large number, 2,042,290 or approximately 77% of messages, contained "re:" or (re). Only 2,238 messages subject lines contained the word reply. One hundred messages from the 2,042,290 messages with any of these reply indicators in the subject line were chosen at random for examination by two human reviewers to determine if they were replies. Both reviewers concluded that 100% of these messages were replies.

There were only a small percentage of messages with indications that they were forwarded. 1059 subject lines contained "FW ", 1854 "FWB", 1370 "FWD", and 4240 had content containing "contains original message". These 8,523 messages represented only 0.3% of messages. One hundred messages with forwarding indicators were chosen at random for examination by two human 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.

Messages often refer to early messages by containing a line with the format similar to: "Name <email 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 contained such strings, representing 45% of the messages sent. One hundred messages with message body text indicating it is a reply were chosen at random for examination by two human reviewers to determine if they were indeed reply messages. Both reviewers concluded that 99 of these 100 messages were replies.

The ">" is the most common indenting used in Usenet messages to signify text quoted from an earlier message. Multiple indenting indicates that it is a quote from a message that it 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 were also examined (e.g. number of lines that started with ">>>" or "> > >"). 1,514,297 or 57% of messages were found to

[1] This approach was taken because unlike information retrieval analysis which typically examines precision we are not interested in the probability of retrievals being relevant.

0-7695-1435-9/02 $17.00 (c) 2002 IEEE

have more than two lines with '>' indentation. More than two lines of indenting was used as a measure of 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. One hundred messages with an indenting measure that suggested they were replies were chosen at random, and were then examined by two human reviewers. The reviewers concluded that 99 of these 100 messages were replies.

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. 1,749,532 messages contained 'references' to other messages, this represented 66% of the study sample. 83% of the 'message-id's contained in these header fields referred to messages posted to the same discussion group within the study period. This proportion was slightly higher if we ignore the first two weeks 87%, or first month of the study 87%. This is not surprising as the average response between the messages that were referenced that were identified in this manner was 1 day, 90% of responses occurred with the first 2 and a half days, 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' representing only 1% of the messages sampled. One hundred messages from the 1,749,532 messages with 'references' were chosen at random for examination by two human reviewers to determine if they were replies. Both reviewers concluded that 100% of these messages were replies.

Table 1. Reply Indicators

Type of Indicator	Number of messages	Percent of Study Sample	Two-Coder Assessment Of Reply Status
Subject Line Reply Indicators	2,042,290	77.0%	100%
Referencing	1,749,532	66.0%	100%
Indentation	1,514,297	57.0%	99%
Words in Content Indicating Reply	1,201,541	45.0%	99%
Forward Indicators	8,523	0.3%	67%

Simply using the existence of any of the indicators described in Table 1 to conclude that a message is a reply is not likely to result in maximum specificity and or sensitivity. All these measures are inter-related, for example, 84% of messages with "re:" or "(re)" in the subject line also have message references, and 98.5% of messages with referencing also have reply strings in the subject line. It was therefore decided to use a formula that combines all these measures in a subtler manner in order to gain improved accuracy. If the subject contained a reply measure (e.g. "re:" or "Reply:") it was considered a reply, because this represents the largest group of messages and is strongly indicative of it being a reply. If a reply string was not found in the subject then messages had to have a score of 0.8 or higher to be considered as a reply. More than two lines with indenting was worth 0.6. "In-reply-to:" was worth 0.5. Having message content that indicated that it was forwarded was worth 0.4. Finally, a message reference was worth 0.3. This formula takes into account the large percentage of messages with a subject line indicating it is a reply and the fact that the vast majority of messages with referencing also have reply strings in the subject. In the end 2,061,179 messages or 78% of the study sample were coded as replies. One hundred messages coded as replies were chosen at random for examination by two human reviewers to determine if they were replies. Both reviewers concluded that 100% of these messages were replies. One hundred messages that were coded as not being replies were chosen at random for examination by two human reviews to determine if they were indeed not replies. Four of these messages were falsely coded, and for 3 messages the status was unclear as even after reading the messages in context it was not possible to tell whether they were replies, so it was concluded that the measure has a specificity of 96%.

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 as a labeled a reply is 99% based on the assumption that if a larger sample of messages was used that the true positives would have dropped below 100% to around 99%. This number is fairly robust because the large majority of messages are replies.

3.3 Identifying Parent Messages

Considering the fact that 16% of messages coded as replies did not have references it was concluded that threading could not be built by a reliance solely on the references contained in the message headers. It was decided to proceed 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. Second, if the parent was not found then a similar step was taken for the "In-Reply-To:" field. 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 backwards sequentially for 14 days and then if needed

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 IEEE

forward half a day. This time frame was chosen because as noted above 99% of responses occurred with the first two weeks. 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 best explained by inaccurate system clocks on client computers).

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 or 73% of reply messages. Interestingly, for those messages with references, it was possible to identify their parent message in 83% of the cases, and this proportion was slightly higher if we ignore the first two weeks, 87%.

One hundred message-reply pairs were extracted at random from the 1,502,991 message-reply pairs contained in the database. Two human coders 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 coders. So the specificity of parent messages 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. Two coders examined all the reply messages from one newsgroup where no parent was found via the formula described above, in order to assess the adequacy of the parent search algorithm. Using this extremely time consuming process, for 67 (40%) of the 170 replies with no parent message identified by the algorithm used it was possible to find the parent message. This suggests 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. It therefore seems safe to conclude that approximately 87% of parent messages were identified. If we consider the probability of falsely having a positive test result as equivalent of stating that a parent message does not exist when it actually does (0.4) then the sensitivity of the algorithm is 92.4%. An examination of the 67 reply messages where a parent was found manually, suggests that two significant improvements to the method used could be made. First, the subject string search needs to take into account the likely depth of the message, in cases where the parent message is not found. For example a message with "re: re: re:" requiring a subject string search should look for messages with "re: re:" in the subject. Second, quoted/indented text in these messages could have been used to examine less indented or un-indented text of potential parent messages. Without implementing these refinements, approximately 87% of the parent messages were identified, a percent that was deemed adequate for the task at hand. However, it may be the case reply messages with unidentified parent messages are significant as they may place a larger cognitive processing strain on the reader who may try to place them in context.

Two measures were computed of the thread depth of each reply message. The first was calculated taking the maximum indenting depth or number of references in the message header. If for example, 3 lines had indenting of '>' and two of indenting of '>>' and one of indenting of '>>>' and one message reference the depth was calculated to be three. If there were 10 references and no indenting then the depth was taken to be 10, etc. The second method for calculating a reply message's depth was by the means of the Oracle SQL extension 'connect by' command, which provided such information based on its reconstruction of discussion threads. For the study 'reply-messages' for which a parent message was found, and Oracle was able to reconstruct threading (1,499,134 messages), the average depth computed by examination of individual messages was 2.36 and the average message depth calculated using Oracle SQL extension 'connect by' command was 2.28. The greatest thread depth based on individual message estimates was 20 while using Oracle SQL 'connect by' command it was 37. These two approaches were moderately correlated when all replies with parent messages were examined (Pearson's Correlation = 0.46, n=1499134). Collectively these findings suggest that an examination of only a single message lacks the context to be truly informative about its message depth. However as only approximately 87% of parent messages were found for the current analysis both approaches can be considered to have limitations.

3.4. Hypothesis Testing

With the demonstration of adequate sensitivity and specificity it is now possible to examine the three hypothesized effects of the information overload induced non-linear feedback loops.

3.4.1. **Hypothesis One: Generating Simpler Responses in Situations of Overloaded Mass Interaction.** Two hypotheses are examined with respect to the generation of simpler responses in situations of overloaded mass

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 IEEE

interaction. These are: 1) there will be a decrease in surrogate measures of complexity of interactive message communication, such as word count, as the size of the interactive group increases although this will approach asymptote; and 2) there will be a decrease in surrogate measures of message complexity (e.g. the number of new words per interactive message) as the number of discussion threads in the newsgroup increases, although this will also approach asymptote. The reason for this hypothesized reduction in message complexity is due to the increased effort required by authors to create such messages. Clearly, there is no absolute measure of message simplicity/complexity although there should be a rough correlation between various message characteristics and the effort required to create and read them. There are number of possible surrogate measures of message complexity, so the first step required is to examine these measures and their appropriateness. The second step is the systematic plotting of these measures of message complexity against the size of the interactive discussion group, as well as against the number of threaded discussion threads. Finally, statistical significance tests are made to determine if any of the trends observed are simply due to chance.

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 were calculated:

- The number of words;
- The number of words on non-threaded lines (no indenting) (new words);
- The number of lines according to the header field;
- The number of lines excluding those of attachments (lines); and
- Number of non-threaded 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 including the concept/s the author is intending to convey, the context of the message in the discourse stream, the complexity of the language required, etc. A more refined measure could perhaps be computed via the use of an algorithm that combined variables computed with information such as sentence length and technical word use, and recognition of the extent to which a message was impacted on by being part of a discussion thread. However, at this stage such analysis is not called for.

The size of interactive discussion group can be determined a number of ways, which in part depends on how the notion of a group is conceived. Is the size of the group the number of subscribers, the number of contributors, the number of messages, the number of contributors to 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. So it makes sense to distinguish posters of responses from others 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. The next issue is time period: should this be for a day, week, month or some other duration? While there is no absolute answer to this question the data plots suggest that for some issues looking at the data from a monthly perspective results in low resolution. On the other hand daily and weekly results appear to be quite similar although there are some quite strong daily fluctuations. This suggests that a weekly measure may be most appropriate. Further, Usenet servers typically store postings for about a week suggesting that looking at a weekly set of data would more closely approximate user interactions with newsgroups. Therefore, for the current purposes, interactive discussion group size will be considered to be equivalent to the number of interactive or reactive posters to a newsgroup over a one-week period.

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 some face validity to the hypotheses under examination.

Figure 2 below, is a scatter plot of the average number of message lines according to header information by the number of posters of threaded messages. The shape of the curve produced by the scatter plots for all the various measures of message complexity and the two measures of size of the interactive discussion group all look similar. The plots displayed the expected relationship between the size of the interactive newsgroup and various surrogate measures of complexity. The scatter plots also show that a standard linear regression cannot be used to describe the untransformed relationship between the measures because of the Zipf [24, 9] like shape of the curve, with a clearly nonlinear relationship. This complicates any effort to provide probability and regression statistics associated with the figures. Further, the curve cannot be

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 IEEE

tested as a standard Zipf/Power curve because the points on the plot represent means rather than frequencies. For this reason multiple transformations of the variables understudy were examined to see if it was possible to produce plots that would allow for the examination of the hypotheses by a linear regression. The transformation approach did not succeed in enabling regression modeling using weekly averages for newsgroups so two alternative approaches were taken. The first approach was simply to divide the distribution of both the number of interactive posters and the number of unique threaded threads into quartiles, and compare the means for the complexity measures to determine if the means decreased as newsgroup activity increased. The second approach was to look for the hypothesized effects by examining individual messages rather than aggregations; this enabled regression modeling using ranks.

Figure 2 Scatter plot of the average number of message lines according to header information by the number of posters of threaded messages.

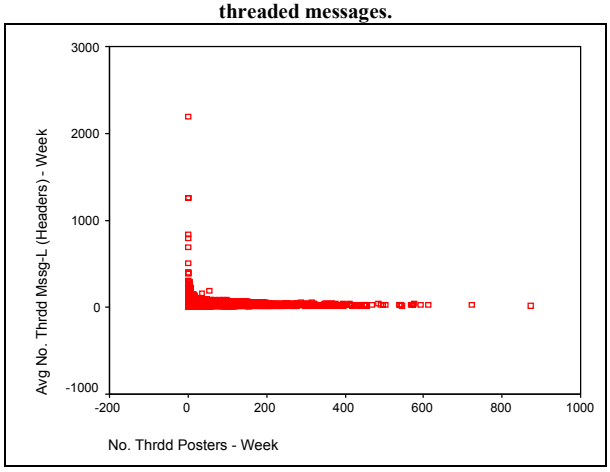


Table 2 below shows the comparison between first and fourth quartiles of group size (number of posters of threaded messages) for various measures of average message complexity. For all measures, the message complexity is reduced as the group size increases and this is the case for all quartiles examined and for both measures of group size (all differences are highly significant).

Unfortunately, while the plots and quartiles data support the model proposed, they do not provide strong evidence. This is because in statistical terms the plots and the tables can be explained in terms of "regression towards the mean", where larger groups have less variance of mean measures [18]. 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 slop as group size increases. Not only that, but the distribution of number of weekly unique threaded threads, and the number interactive posters, is such that the sample has more small groups. As a result the forth quartile for number of unique posters was 45 and up, and for number of unique threaded threads was 59 and up. On the other hand, the comparison between the third and forth quartiles were still significant and the means are decreasing. Of course these findings should be understood in the context of the other findings presented in this paper, including the evidence provided by regression modeling using the ranks of individual message complexity as the independent variable.

Table 2. Average Message Complexity Measures for 1st and 4th Quartile of Distribution of Posters of Threaded Messages

	Number of Threaded Posters in Quartiles	N	Mean	Std. Deviation	Std. Error Mean
Avg No. Thrdd Mssg-L (Headers) - Week	First Quartile	3164	36.143526	62.463994	1.110482
	Fourth Quartile	3072	30.511169	8.236052	.148593
Avg No. Thrdd Mssg-L (Content) - Week	First Quartile	3164	34.392979	61.888017	1.100242
	Fourth Quartile	3072	28.880758	8.039659	.145053
Avg No. Thrdd New Mssg-L (Content) - Week	First Quartile	3164	20.485091	52.030473	.924995
	Fourth Quartile	3072	16.490232	3.928647	7.09E-02
Avg No. Thrdd Mssg-W (Content) - Week	First Quartile	3164	208.6809	353.416610	6.283024
	Fourth Quartile	3072	182.9633	55.013664	.992567
Avg No. Thrdd New Mssg-W - Week	First Quartile	3164	126.2465	275.825911	4.903620
	Fourth Quartile	3072	106.7022	29.033123	.523821

To perform regression modeling the 1.5 million threaded messages were ranked according to their comparative message complexity. As shown above, the most useful measure of message complexity was the average number of message lines calculated by the posters client newsreader. Information about the newsgroup activity during the study week messages were posted was then matched to individual messages. Using this approach it is possible to see if the number of posters and or number of interactive threads is at all predictive of message complexity without the concern of 'regression to the mean'. This approach also allowed for factors such as newsgroup type (e.g. Comp. or Misc.) and message crossposting to be taken into account. Unfortunately it also results in a loss of variance and predictive power. In other words, the ranking approach with transformation enabled a regression to be run, although this necessitated approach meant that the model's explanatory power was not great. Instead, the aim was to see if further support could be found for the notion that group activity related to message complexity. 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) as did the type of newsgroup messages were posted to, the extent of crossposting (messages that were crossposted were longer on average), and the messages' position / depth in a discussion thread (deeper messages were longer overall) ($F=14836.24$, $df= 1499124$, $p < 0.0000$).

3.4.2. Hypothesis Two: Failing To Respond Or Attend To Certain Messages. When users are confronted with overloaded mass interaction it was hypothesized above that they are more likely to fail to respond and / or attend to the messages that are more onerous to process. It follows that simpler messages will be more likely to seed (start) new discussion threads than complex messages in overloaded discourse. There were a total of 593,019 messages that could be considered true unambiguous broadcast or one-way messages. Of this sample of one-way messages 255,697 were found to have initiated (seeded) discussion within their newsgroup during the study period. Tables 4 and 5 below compare the means for various message complexity measures and average newsgroup crossposting between broadcast messages that seeded and those did not seed further discussion.

Table 3. Means Table of
Average Broadcast Message Complexity & Discourse Seeding

	ISSEED	N	Mean	Std. Deviation	Std. Error Mean
Client Header Estimate of Number of Lines	Not Seed	337322	62.289442	645.656535	1.111679
	Is Seed	255697	24.501574	225.340135	.445631
Number of Message Lines	Not Seed	337322	53.525367	154.553286	.266107
	Is Seed	255697	22.526060	51.012182	.100881
Number of Non-Indented Message Lines	Not Seed	337322	53.223232	154.410399	.265867
	Is Seed	255697	22.336656	50.751880	.100367
Number of Message Words	Not Seed	337322	318.5668	958.131937	1.649692
	Is Seed	255697	144.8529	340.536468	.673443
Number of Non-Idented Message Lines	Not Seed	337322	316.6876	956.970562	1.647693
	Is Seed	255697	143.6009	338.455536	.669328
Message Crossposting	Not Seed	337322	1.492912	1.481911	2.55E-03
	Is Seed	255697	1.281239	1.020754	2.02E-03

As predicted, table 3 show that, on average, broadcast messages that seed discourse are smaller/shorter than those that fail to seed discourse (all differences between highly statistically significant). They are also less likely to have been crossposted. Of course the analysis in table 3 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. For example, the "comp." newsgroups, which are about various computer issues, are very active and being focused on technology would perhaps contain shorter messages than for example the smaller recreational ("rec.") discussion groups. Fortunately, unlike the examination of simpler message generation in overloaded situations a number of these issues can be examined directly by regression, and there is no problem of 'regression to the mean'. This is because the binary outcome of either seeding or not seeding further discourse allows for the use of logistic regression techniques, whose underlying mathematical model is nonlinear. By the use of this approach it was possible to assess, when controlling for factors such as newsgroup activity and newsgroup topic, if message size did indeed relate to seeding new discourse. More specifically the approach taken was to put all the possible explanatory variables in the model, and then via backward elimination, remove items that did not enhance explanatory power.

To understand the relationship between various variables and likelihood that a message would seed discourse, a very large number of potential logistic regression models were examined. The outcome of regression modeling was that the following factors are all predictors of a one-way messages seeding new discourse, after controlling for other explanatory variables:

- The overall activity of the newsgroup (measured by the number of messages posted per week);
- All the measures of message complexity (e.g. number of words);
- Newsgroup type (e.g. 'talk' and 'misc'); and
- Moderation status (using a variety of approaches including newsgroup name and Newsgroups Information Center descriptions).

As one would expect, one-way messages posted to larger more active groups were more likely to be responded to. Further, all measure of message complexity was found to negatively correlate with seeding discourse (i.e. smaller messages were more likely to seed discourse) and the client header calculation of message length, which is influenced by attachment lengths, was found to be the best predictor. Posting a one-way message to the 'comp' or 'sci' newsgroups resulted in a greater chance of receiving a reply than posting to the other newsgroups, and finally moderation reduced the chances of receiving a reply. The logistic model that appears to best describe the dynamics of discourse seeding was able to predict 63.57% of the cases with a Wald χ^2 of 56559.408, $p > 0.0001$. The findings of the logistic regression modeling argue strongly for the conclusion that smaller messages are more likely to generate ongoing discourse. Of course, this is not to say that larger messages are more likely to be of a type, such as FAQs, summations etc., which are not designed to lend themselves to discourse.

3.4.3. Hypothesis Three: Ending Active Participation. The final hypothesis to be examined is that a higher proportion of active users will end their active participation in larger more overloaded discussion groups. This phenomenon was hypothesized to occur because disengagement is one strategy users can adopt to cope with overloaded discourse. It follows, then, that on average at AvMaxCL the larger the number of individuals involved in discourse the less stable the population of active participants. The reason for

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 IEEE

examining proportions as opposed to total numbers is simply that larger groups could potentially have a greater number of individuals disengage from discourse. To assess the validity of this hypothesis the scale at which stability is to be examined needs to be decided upon. Like all the measures examined thus far, because of the newness of this research, there is no obvious standard based upon previous research to choose from. An examination of the Usenet data shows that many active newsgroup users do not post every week suggesting that the measure of stability 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. So, for the purposes of this study, proportional membership is the percent of posters per-newsgroup per-month, who also posted to the next study month. This allowed for the examination of user stability over a 5-month period.

0-7695-1435-9/02 $17.00 (c) 2002 IEEE

Figure 3. Proportional Usenet Newsgroup Stability

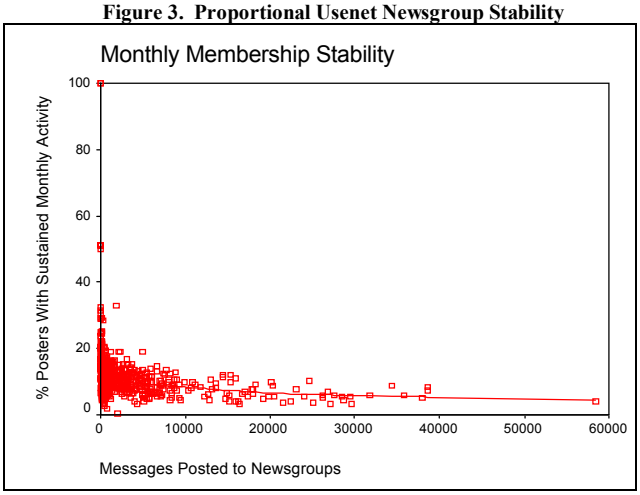


Figure 3 displays the number of messages posted to the 578 newsgroups that were active during the first 5 months of the study. On average only 11.5% of posters sent messages 2 months in a row. Because of the constraints imposed by the proportionality of the stability measure (zero to one hundred) it seems reasonable in this case to also plot on Figure 3 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 -.43 ($p < .000$, n=565). 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 2957 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 ($p < .000$, n=192), highlighting the strength of this finding. If the outliers and moderate skew of the proportional stability measure is ignored and linear regression modeling is used on the full sample then the number of posters (the best predictor, the more posters the lower the stability), moderation status (moderation increases stability), average newsgroup message crossposting (more crossposts results in greater stability) and newsgroup type are all found to predict membership stability (R^2=.24, F=22.1, df=556, P < .001). Such regression modeling is not however ideal because the outliers result in a non-normal distribution of regression residuals. It was therefore decided to run a regression on the top third of the sample as this removes the outliers that probably result from small group size, and results in a normal distribution of regression residuals. The outcome was similar although the discriminatory power of the regression was greatly improved (R^2=.43, F=17.7, df=183, P < .001), with group size being the best predictor of stability.

4. Discussion and Conclusions

The emergence of mass interaction has presented new opportunities to learn about and understand human communication. The availability and persistence of such communications, and the scale at which it operates allows us to explore various system effects on group discourse. At this point in time empirical research into the *systemic* nature of the patterning of social relationships in cyberspace has, despite its importance, been relatively rare.

Work which use a systems approach to examine internet based group communication, such as: the modeling of free riding using the Napster like Gnutella network [3]; modeling the inter-relationship between homepages [2]; exploring the self-organizing nature of email lists [7]; and showing the World Wide Web to be structured like a small world network [1]; have been undertaken in the last five years. The work described in this paper is the first to explore empirically the impact of systems effects in Usenet discourse. Perhaps more importantly, the hypothesized effects were generated from various theoretical works [13-16] that together suggest a research program into the nature of mass-interaction dynamics and its impact on CMC-technology use. The research program is based upon the existence of cognitive processing constraints that result in non-linear feedback loops that in turn impact on mass interaction discourse structures. As shown in this paper, these impacts can be examined empirically.

Overall, the results strongly support the assertion that individual 'information overload' coping strategies have an observable impact on mass-interaction discourse dynamics. Clear evidence was found for the hypotheses: that users are more likely to respond to simpler messages in overloaded mass interaction; and that users are more likely to end active participation as the overloading of mass-interaction increases. Evidence was also found for the hypothesis that users are more likely to generate simpler responses as the overloading of mass-interaction increases.

The research program that guided this research suggests a number of other ways to further test and improve upon this work. Possible future areas include: examining the relationship between message error rates and group size; examining the findings using improved thread reconstruction techniques; and taking into account message types (e.g. FAQ's, Announcements, Questions,

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 IEEE

etc.). The research program holds that the way non-linear feedback loops impact on mass interaction relates to the CMC-technology under investigation. This is because different CMC-technologies will have different *typical* message system characteristics. Therefore, the point at which a user population's interactions will typically result in information overload will relate to the CMC-tool used. It follows then that the findings presented in this paper pave the way for comparative analysis of the usability of various computer-mediated communication technologies. In this regard, related research shows that the dynamics of email list mass interaction is radically different [16].

It follows from the above that the system dynamics approach allows for the examination of computer mediated communication technologies in terms of group-level usability. 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" [6]. This supports the use of usability laboratories, and ethnographic methods, which can put user behavior in context. While, not discounting the value of these approaches, using the methodology presented in this paper for comparitive purposes represents an alternative approach. This is because it potentially allows us to see and compare the normal range of user interaction dynamics for differences types of CMC-technologies.

5. References

1. Adamic, L. A. (1999). The Small World Web. Proceedings of the European Conf on Digital Libraries.
2. Adamic, L. A. and E. Adar (2000). "Friends and neighbors on the Web".
3. Adar, E. and B. A. Huberman (2000). "Free Riding on Gnutella." First Monday 10(5).
4. Armitage, P. and G. Berry (1987). Statistical Methods in Medical Research. Oxford, Blackwell Scientific Publications.
5. Berger, C. R., S. W. Knowlton, et al. (1996). "The hierarchy principle in strategic communication." Communication Theory 6(2): 111-142.
6. Bevan, N and M Macleod, (1994). Usability Assessment and Measurement. In: *The Management and Measurement of Software Quality,* (Ed. M. Kelly), Ashgate Technical/Gower Press..
7. Ekeblad, E. (1999). The emergence and decay of multilogue: Self regulation of a scholarly mailnglist. European Association for Research on Learning and Instruction (EARLI), Sweden.
8. Finholt, T. and L. S. Sproull (1990). "Electronic groups at work." Organization Science 1(1): 41-64.
9. Gunther, R., Shapiro, L., and Wagner P., (1996). "Zipf's law and the effect of ranking on probability distributions." International Journal of Theoretical Physics 35(2): 395-417.
10. Herring, S. C. (1999). Interactional coherence in CMC. Proceedings of the 32nd Hawaii International Conference on System Sciences, Hawaii, IEEE.
11. Hiltz, S. R. and M. Turoff (1985). "Structuring computer-mediated communication systems to avoid information overload." Communications of the ACM 28(7 July).
12. Ingelfinger, J., F. Mosteller, et al. (1987). Biostatistics. New York, Macmillan.
13. Jones Q. (1997). "Virtual-communities, virtual-settlements & cyber-archaeology: A theoretical outline." Journal of Computer Mediated Communication **3**(3).
14. Jones Q. and Rafaeli S. (2000b). What do virtual 'Tells' tell? Placing cybersociety research into a hierarchy of social explanation. 33rd Hawaii International Conference on System Sciences, (Hawaii 2000), Hawaii, IEEE Press.
15. Jones Q., Rafaeli S. (2000d). "Time to Split, Virtually: 'Discourse Architecture' and 'Community Building' as means to Creating Vibrant Virtual Publics." Electronic Markets, The International Journal of Electronic Commerce and Business Media. 10, (4): 214-223.
16. Jones Q. (2001). "The Boundaries of Virtual Communities: From virtual settlements to the discourse dynamics of virtual publics." PhD Thesis, Graduate School of Business, University of Haifa, Israel.
17. Lewis, D. D. and K. A. Knowles (1997). "Threading electronic mail: A preliminary study." Information Processing and Management 33(2): 209-217.
18. Newell, D. and Simpson. J (1990). "Regression to the Mean." The Medical Journal of Australia **153**(August 6): 166-168.
19. Rafaeli, S. and Sudweeks F., (1997). "Networked Interactivity." Journal of Computer Mediated Communication 2(4).
20. Rheingold, H. (1993). The virtual community: Homesteading on the electronic frontier. Reading, MA, Addison-Wesley.
21. Sastry M. and Sterman J. (1992) Desert Island Dynamics: An Annotated Survey of the Essential System Dynamics Literature. Online June 2001 http://web.mit.edu/jsterman/www/DID.html
22. Sproull, L. and S. Faraj (1997). Atheism, sex and databases: The Net as a social technology. Culture of the Internet. S. Kiesler. Mahwah, NJ, Lawrence Erlbaum Assoc, Inc.
23. Whittaker, S., L. Terveen, et al. (1998). The dynamics of mass interaction. CSCW 98, Seattle, ACM Press.
24. Zipf, G. K. (1949). Human behaviour and the principle of least effort. Cambridge, MA.

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 **IEEE**

0-7695-1435-9/02 $17.00 (c) 2002 IEEE

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 IEEE