The Commons: A Toolkit for the Quantitative Analysis of Factors in the Success or Failure of Electronic Societies

Joshua Berman
Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052 USA
+1 404 786 1698
berman@cc.gatech.edu

Steven M. Drucker
Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052 USA
+1 425 703 5880
sdrucker@microsoft.com

Peter Kollock
Department of Sociology
University of California at Los Angeles
Los Angeles, CA 90095-1551
USA
+1 310 825 1313
kollock@ucla.edu

ABSTRACT
In this paper, we introduce The Commons, which is a novel methodology and toolkit to aid in the design and study of online societies. We first discuss the foundations which led to its inception, and explain the design decisions made along the way to its completion. Next, we describe the system itself. We then describe and analyze the results of a user study, which shows that this methodology is valid, and that the toolkit produces results comparable to the existing social science literature. In this study we examine a particular form of online communication -- synchronous text chat - and report on its effects in increasing trust and cooperation in online interactions. It is a common intuition that communication leads to more cooperative online relationships but intuition is not proof, nor does intuition provide a precise mapping of the effects and dynamics of various forms of communication on cooperation. Next, we will examine the usefulness of The Commons in the design of successful electronic societies by exploring its advantages over those methods most commonly used today. Finally, we discuss the importance of this platform for continued experimental research in both the social science and computer-human interaction research communities.

Keywords
Electronic Societies, Virtual Communities, Social Interfaces, Social Computing, Computer-Mediated Communication, Community Computing, Social Dilemmas

INTRODUCTION
What does it mean for an electronic society to be successful? Many researchers over the past two decades have hypothesized about what it takes, but very few have precisely defined what it means.

Some researchers have simply used 'community' as a descriptive measure, so that a successful online society would have more 'community'. Affordances and societal choices which led to its success would be said to have 'built community'. However, this approach does not help designers, developers or maintainers of electronic societies. They ask how this measure can be compared across different societies. They wonder how to tell whether their work is increasing or decreasing this abstract factor.

There has been some attempt to measure success in terms of "retention factor"[4] -- a survival-of-the fittest approach, where success can be measured by user count. Of course, this means that the effect of new affordances and societal changes can not be measured, and that advertising and market pressure must guide even researchers' view of how to create successful online society.

Others have argued that that success can be measured in terms of "village like-quality"[3] or how well they "resemble 'real life' communities"[19]. This method begs the question. What makes a successful village or successful 'real life' community? It is the very fact that these are questions that are applicable to non-electronic societies which led to our solution to the problem. If the same standards for success of society apply online and offline, is there a metric for the success of societies in general? Could this metric be quantified, and used to study electronic societies specifically?

One such metric has already been developed. Across a variety of disciplines in the social sciences, a key way of conceptualizing and measuring the success of a relationship or group is to look at the level of trust and cooperation that exists. In fact, a well-defined set of models and methods has been developed to test the existence and dynamics of cooperation in social relations [9,12,14].

We have created a system that capitalizes on these methods to create a rich and extensible toolkit that can be used to better understand and develop successful electronic societies.

PREVIOUS AND RELATED WORK

The research area we draw upon from the social sciences is the cross-disciplinary work on *social dilemmas*. Social dilemmas are those situations in which an individually reasonable decision leads to collective disaster, that is, a situation in which everyone is worse off than they might have been otherwise. Models of social dilemmas capture this tension between individual and collective outcomes, and can therefore be used as a very powerful and broadly applicable probe to assess the level of cooperation and trust in a group. Since the 1950s, a large research literature has developed in this area (for reviews, see [9,12,14]).

However, there has been surprisingly little current work that has applied social dilemma models to online interaction and electronic societies. A recent exception is Rocca's work that examined the difference between electronic and face-to-face communication [16]. In Rocca's research, a specialized system was built, and a specific experiment was run, to show the difference between online and face-to-face communities. What we have completed is instead a generalizable toolkit and methodology capable of capturing not only these basic differences, but differences created by interface or social factors within electronic societies themselves. The toolkit created, which we have named *The Commons*, allows and facilitates the usage of social dilemma paradigms to study electronic societies, as shown in Figure 1.

THE COMMONS

The Commons toolkit was created using the V-Worlds Platform, a robust and extensible framework for distributed persistent objects [18]. The V-Worlds system, specifically designed to facilitate the development of shared virtual environments, offered many advantages, namely, automatic communication between multiple clients, easy programming via scripting, convenient interface prototyping via Dynamic HTML, and integration with the World Wide Web.

The fact that the V-Worlds Platform was designed specifically for the creation of shared virtual environments aided in the development process. In particular V-Worlds maintains the notion that multiple clients can be connected to a virtual room and that all clients automatically receive updates for any state changes to objects occurring within the room. Hence, much of the communication infrastructure was already completed, and creating the specialized types of communication for The Commons was often a matter of just a few lines of code.

In addition, because of the modular nature of the platform, and the ability to modify objects even at run time, different behaviors could be easily prototyped. This created the ability to quickly and easily modify an existing environment to include The Commons, and to change experimental conditions with little change to the rest of the interface and user experience.

Also, V-Worlds was specifically designed to allow world creation and modification in Active Scripting languages. This allowed us to create a toolkit which could be used easily by members of the sociology, psychology and design community without the need for dedicated developers. For example, to change the reward payoff for a social dilemma game, an experimenter must simply change a single variable in a JavaScript textfile. No knowledge of programming is required.

Finally, the V-Worlds architecture, integrated in design and practice with Microsoft Internet Explorer, allowed us to create tools which could be run on the World Wide Web, allowing The Commons to run experiments efficiently and inexpensively which were not previously possible. Examples of these types of experiments would be those done over a great geographical distance, an extended period of time, or those wishing to use the large user population of the World Wide Web. The potential advantages of this platform for experimental research are detailed at the end of this paper.

A set of interface elements, along with a related set of internal states and methods, were added to the V-Worlds system, to create a toolkit for the modeling of social dilemmas. For our initial studies, we chose to focus on an iterated Prisoner's Dilemma Game. The game is iterated in that partners play it repeatedly with each other. This allows relationship to build over the course of the experiment, and effects to be recognizable which might only occur over time. For our first study, we focus on the dyadic (2-person) version of the game, although future studies will examine the multiple person (termed N-Person) versions of the game as well. We also chose to use a continuous version of the game: in the classic Prisoner's Dilemma, subjects are given the dichotomous choice of cooperating completely or defecting completely. This is an unrealistic and restrictive assumption given that interaction usually involves degrees of cooperation [10,22]. For this reason, we permit a range of cooperation. The rules of this game are given in figure 2.

Thus, the situation has the structure of a Prisoner's Dilemma: The greatest possible return comes from keeping all of one's points while one's partner contributes all 10 points (a return of 30 points -- the 10 original points plus the 20 points from the partner's doubled contribution). However, if both actors follow this strategy each will end up with only 10 points (having contributed none to each other) rather than the 20 points each could receive if each contributed all their points.

Although this game was chosen as an initial social dilemma, other games can be easily modeled with the same toolkit by making simple scripting changes. However, this choice of a default dilemma model has several benefits. As opposed to discrete dilemmas where the only options available to players are to cooperate or defect, the continuous nature of this game allows experimenters to see more finely grained differences in levels of cooperation. Also, the simplicity of the game, as evidenced by the brevity of the instructions,

allows the games to be run quickly without prior user training. Finally, the dyadic nature of the game allows experiments to be run without the necessity of coordinating or compensating large groups as in the polyadic cases. This also means that experimenters can control for many confounding variables present in group cases, concentrating instead on those variables the experimenters wish to study.

Using this game, a module was created for V-Worlds that would provide all of the necessary affordances not only for the play of the game, but for the recording of any data that would be relevant to or of interest to the experimenter. Again in this process every effort was made to preserve the flexibility of the toolkit, allowing for changes in data presentation and recording to be made simply and easily in Active Scripting languages.

Once the general toolkit was created and tested, scripts were created in order to test the validity of this approach by running a simple but important experiment. Using the Commons toolkit, we designed and implemented an experiment to study the effects of communication on levels of cooperation in online interactions.

Specifically, we examine a particular form of online communication – synchronous text chat – and report on its effects in increasing trust and cooperation in online interactions. It is a common intuition that communication leads to more cooperative online relationships – electronic society designers and maintainers are well aware that lack of communication ability can severely undermine attempts to create a successful community [1,20]. However, intuition is not proof, nor does intuition provide a precise mapping of the effects and dynamics of various forms of communication on cooperation. What is called for is a series of experimental studies that determine the effects of different kinds of information flows and how those effects change across time and across different interfaces. We begin this research agenda here.

THE VALIDATION STUDY

Introduction

In this study, subjects played an iterated dyadic continuous prisoner's dilemma for seventy rounds. They did not meet each other face-to-face before, during, or after the game. Approximately half of the subjects who played were allowed to send text messages to each other (chat) during the game. The other half had no communication beyond what the game itself provided. Data was gathered on the contribution made by each player at each round, as well as data about chat message counts and contents.

This data was then analyzed to determine whether there was a significant difference in the level of trust and cooperation between the two groups of subjects. Based on the previously cited work in the creation of electronic communities, we hypothesized that there would indeed be such a difference, with the group able to chat having the higher level of trust and cooperation. Therefore, we were interested in showing that The Commons could demonstrate this widely assumed fact to be true, and how descriptive it would be of its effect.

Further, we used this experiment to check the validity of the approach in general by comparing contribution rates in the group without chat with other similar studies from social dilemma research which were not performed online.

Subjects

The subjects were 20[1] adults aged 26 to 54 years old (average=41.05). Approximately half of the subjects were men (n=12) and half of the subjects were women (n=8). They were recruited from a non-employee Microsoft subject pool, and only subjects with intermediate Internet skills and some experience with Microsoft Internet Explorer were selected to participate. In addition to the compensation related to performance in this specific experiment as will be discussed later, subjects received a gratuity of one Microsoft software package of their choice for participation.

Procedure

The subjects were randomly assigned to two groups. All factors were held constant between the two groups with the exception of the fact that one group (the chat condition) was given instructions and affordances for sending and receiving text messages with their partner during the experiment.

Each participant, isolated from all other participants throughout the experiment, was given a preliminary briefing that their performance in the game would affect their reward upon leaving. Informed consent was gathered at this time. They were then given semi-random three digit user numbers which were used as monikers for the rest of the experiment to control for any effects real names might have created.

Each participant then read a series of instructions using HTML Web Pages. These instructions explained the rules of the game and that it would continue for "approximately 100 rounds". In addition, these instructions were written to reinforce that a subject's success in the game would affect their compensation, and included a test of understanding, which the subjects were required to complete correctly before continuing.

Once both the subject and their partner had completed the instructions and test of understanding, they began to play the game using the interface shown in Figure 3.

During the game, each player's contributions, as well as any chat interaction, was recorded for further study. Play continued without interruption for 70 rounds, at which time the game was ended without warning to avoid end-game

[1] An additional pair of subjects were dropped from the analyses when it became clear that they did not understand the experimental instructions

effects [2]. Each participant then completed an electronic questionnaire containing background and impression data, as well as other measures of interest to the experimenters. Next, each subject was interviewed and asked open-ended questions about their experience. Each was then given extra compensation in the form of gift certificates based on their final score in the game, thanked for their participation, and given their software gratuity.

Results

Analysis of the post-experimental questionnaire provides evidence that the subjects were motivated to earn as much as possible (mean = 5.50; seven point scale), rated the instructions they received as very clear (mean = 6.89), and enjoyed participating in the experiment (mean = 5.00). None of these scores differed significantly between the two experimental conditions.

At the conclusion of the experiment, the data was analyzed for both group differences and trends with time. Therefore there was more than one hypothesis being tested. Due to the small sample size, it was assumed that we would not reach levels of statistical certainty. However, the results, as shown in figures 3 and 4, were clearly compelling, and statistical analyses were performed to determine whether the differences both between and within conditions could be shown significant. Statistics for pairs of participants were averaged to account for the dependency between partners.

Descriptive statistics showed clear differences between the groups, henceforth labeled as chat and no-chat. The means were 8.10 for chat and 4.38 for no-chat. Medians were 10 and 3.25 respectively. It was also noted that the variance was much higher for the condition without chat, 17.14 as opposed to 8.84. Due to the small number of participants, this difference approached, but did not reach, signficance,

$F(1,8)=3.00$, $p\approx.12$. However, using more appropriate statistical approaches, several results could be shown to a level of proof.

The data seemed to suggest that cooperation decreased over time in the no-chat condition, while remaining stable in the chat condition. To test this, repeated measures anova tests were performed to test for the main effect of blocks of rounds across time. For these tests, rounds were considered in fourteen blocks of five. For the no chat condition, round block had an effect of significance, $F(13,52) = 3.06$, $p<.01$. For the chat condition, there was no significant effect of round, $F(13,52) = 1.06$, $p\approx.41$. Combining these results with the graphs of figures 4 and 5 showed these hypotheses to be true.

Finally, although the overall between-subjects difference was not strongly statistically significant, the interaction effect between condition and round was. Precisely, the effect of chat on cooperation rates as compared to that of no-chat differ across the trials, with chat having an increasing relative effect over time. Another way of stating this is that the difference between the two conditions increases over time, $F(13,104) = 2.19$, $p<.02$.

These results showed that The Commons was not only able to show the difference in trust and cooperation which communication provided, but was able to prove it to a higher bar than was needed, namely statistical significance.

OTHER FINDINGS

Another key set of results comes from the post-experimental questionnaire. Subjects were presented with the four possible extreme outcomes of the game (mutual cooperation, mutual defection, exploitation of partner, exploitation of self), and asked to rate the desirability of each outcome on a 7-point scale.

These questions are one way of assessing the subject's *subjective transformation* [7] of the game, that is, the subjective ranking that a person assigns to an outcome, which may, of course, be different than the objective payoff.[2]

Figure 6 displays the mean rating for the four outcomes for each experimental condition. Note that in both conditions, mutual cooperation is ranked more highly than exploitation of partner, even though exploitation of partner (I contribute nothing while my partner contributes all 10 points) results in the greater number of points. Note that subjects in the chat condition compared to the no-chat condition are much more dissatisfied with an outcome of mutual defection or an outcome of successfully exploiting one's partner. In other words, the subjects in the chat condition have a greater

[2] This method of assessing transformations follows [7] and [20]. Subjects' answers to questionnaire items such as these have been shown to be predictive of their actual choice behavior in games involving money [20].

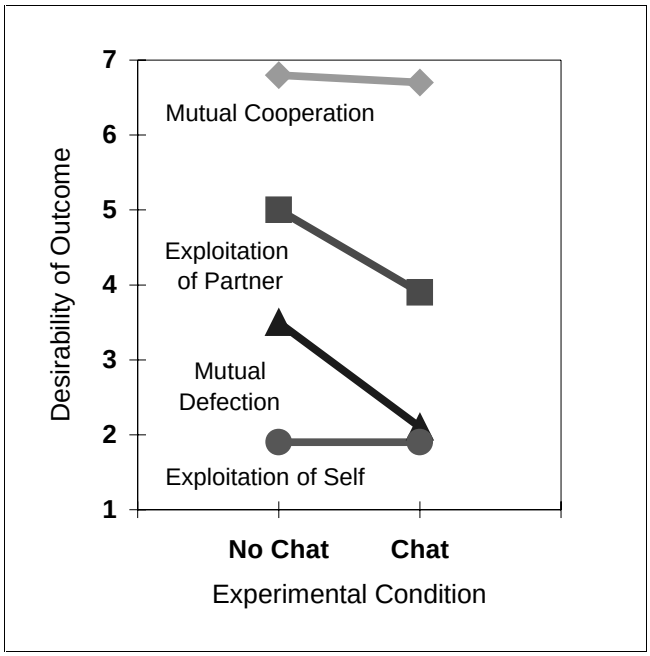

Figure 6. Effect of chat on subjective preference structure.

relative preference for mutual cooperation.

Finally, in the post-experimental questionnaire, the subjects were also asked to evaluate their partner on 16 bipolar scales.[3] Interestingly, subjects in the chat conditions rated their partner more favorably on *all* of these dimensions. A number of these differences approach or reach statistical significance: Subjects in the chat condition on average rated their partner as more helpful than subjects in the no-chat condition ($p = .002$), more flexible ($p = .056$), more rational ($p = .068$), and more honest ($p = .088$).

Collectively, these results demonstrate the manifold effects of synchronous text chat on online interaction. This initial experiment also demonstrates the usefulness of the Commons toolkit for rapidly designing and implementing experimental studies that can be used for both theoretical research and the practical evaluation of different designs for online communities.

VALIDITY CONCLUSIONS

This experiment was also conducted as a validation of The Commons methodology and software, and this toolkit has passed a two-pronged test for validity. First, these results are comparable to previous studies using the same game with other subject populations in other settings [11]. This allows its results to rely on the healthy and growing field of community research in the Social Sciences. Second, the toolkit is able to discover differences in levels of trust and cooperation due to changes in interface. Therefore, the study has shown that the methodology of The Commons is valid, and the toolkit implementing it is correct. Now we turn our discussion to its usefulness.

DISCUSSION

Currently, when evaluating features for online communities, two methods are primarily used. As we will see, although reasonable, both of these methods have severe drawbacks. With the commons, we have provided a novel tool which can not only act as its own method for this task, but can supplement these other methods to increase their effectiveness and minimize their drawbacks.

The first method used for evaluating and choosing features is comparative history. This involves a community designer looking at existing successful communities and trying to take features from them which they believe have helped lead to the community's success. An example would be a designer, upon seeing that the Stacy Horn's ECHO community uses real world user names [6], and deciding that ECHO is successful, decides that real world user names are thus important for community building.

This method has many obvious disadvantages. First and foremost, there is an important inductive step which, instead of being examined and proven, is left to the 'gut feeling' of the designer. In this step, a determination is made about which features of the almost limitless feature list in an electronic society are having positive effects on the society itself. Even if the designer guesses correctly, there is no way for them to test for themselves or prove to others that their choice is correct. This can be shown in our ECHO example. It is possible that the success of ECHO is not helped, indeed it may be hindered by the use of real world user names. This negative effect may be obscured by the fact it is run by Stacy Horn, which may or may not be an advantage, but may vary the success of the community so widely that the presumably smaller effect of real world user names is hidden.

With The Commons, the designer's hypothesis could be tested. It would be possible to run The Commons using real world user names (or some variation) and aliases, and compare the scores. The designer would then not only know, but also be able to prove that real world user names have a positive or negative effect on community building. Of course, as will be discussed shortly, it is not only single features that can be compared using The Commons. Synergistic effects are also measurable using this toolkit.

The other approach currently used in choosing features is intuition and elegance. This includes choosing features from fictional works, and novel features, simply because they appeal aesthetically or 'should work'. An example of

[3] Subjects were asked to indicate their overall impression of their partner on the following scales: intelligent-unintelligent, dishonest-honest, fair-unfair, unlikable-likable, rational-nonrational, untrustworthy-trustworthy, powerful-weak, kind-unkind, unfriendly-friendly, rigid-flexible, aggressive-unaggressive, helpful-unhelpful, cooperative-competitive, cold-warm, sincere-manipulative, selfish-unselfish.

this would be a community builder including a live video of each user's face because 'people use facial expressions to communicate all the time'.

Of course, this approach also has disadvantages. Before the community is built, and, as we showed in the first case, even after it is built, it is impossible to tell whether this feature is useful for community building.

The Commons again can help to solve this dilemma. Simply by running an experiment where the presence of this affordance is the independent variable, this argument can be solved. This can also help to develop novel affordances and community features, as it is now possible to test them without building an entire community.

Of course, it is not necessarily single factors which are combined to create a successful electronic society. It is widely believed that more synergistic or holistic views are also important. The creation of an online society may be more like making a good recipe than constructing a sound building, where it is the interplay between the ingredients which lead to a sum greater than its constituent parts. The Commons allows for those taking a more synergistic view of community creation to perform experiments as well, in turn providing them with useful data.

For example, although it might be shown that the use of real world user names do in fact help in creating a successful online society, it may be true that this effect changes in degree or even direction when combined with other features. Thus it may occur that when combined with other features such as those found in support groups for socially sensitive diseases such as alcoholism, they have the opposite effect. It would be possible using The Commons to use an actual community confronting these issues to test this hypothesis. As stated earlier, the V-Worlds architecture gives The Commons a modular design which is compatible with modern operating systems and the World Wide Web. Therefore, using the actual community, be it web-based or not, The Commons could be run in either the same browsing window or on the same computer, connecting the same participants. If the community used pseudonyms, then The Commons could be easily configured to transmit the users' names to their partners, otherwise, it could be used to mask their names by interacting with the community software. Then, subjects could be run in a condition with and without real world user names. The results of this subject could aid the designers or developers in their decisions about this compound effect.

As a final comment, it is important to point out the significance of this flexible platform as a general tool for experiments in the social sciences and the human-computer interaction field in general. The creation of a "virtual lab" that links subjects across the web represents a key advance in the way social science experiments are conducted, with implications that extend well beyond this initial research application. A Web-based lab addresses a number of limitations in current research [13]:

1. *Research on large groups*. The ability for hundreds or even thousands of subjects to participate simultaneously opens up for the first time experimental research on collective action in large groups.
2. *Increased sample size*. The ability to recruit from a much broader population, and to run hundreds of subjects simultaneously from a single Web-server, also removes current constraints on sample size.
3. *Reduced selection bias*. A Web-based laboratory does not need to be located on a college campus, or even in the same country as the experimenter.
4. *Experimenter effects*. A Web-based lab allows subjects to participate from their own homes and offices, free from surveillance by the experimenter. Social pressures against selfish behavior can then be manipulated, free from contamination by the uncontrolled effects of participation in a physically constrained space controlled by an authority figure.

CONCLUSIONS

The Commons is a methodology and toolkit to aid in the design and study of online societies. We have discussed the foundations which led to the idea, and explained the design decisions made along the way to its completion. We have shown that this methodology is valid, and that the toolkit produces results comparable to the existing social science literature, by performing and analyzing the results from a user study. Finally, we have examined its usefulness in the design of successful electronic societies by showing its advantage over those methods most often used today.

ACKNOWLEDGMENTS

[WITHHELD FOR REVIEW]

REFERENCES

1. Abela, A., and Sacconaghi, A., Value Exchange: The Secret of Building Customer Relationships On Line. *The McKinsey Quarterly 1997*, Number 2
2. Axelrod, R. *The Evolution of Cooperation*. Basic Books, New York, 1984
3. Coate, J. Cyberspace Innkeeping: Building Online Community. Available at http://www.gopher.well.sf.ca.us:70/0/Community/Innkeeping
4. Glaser, M. Building Online Communities. *New Media*, March 3, 1997
5. Hardin, G., The Tragedy of The Commons. *Science*. 162:1243-48, 1968
6. Horn, S. *Cyberville*. Warner Books, Inc., New York, 1988
7. Kelly, H. *Interpersonal Relations: A Theory of Interdependence*. Wiley and Sons, New York, 1978
8. Kiesler, S., Sproull, L. and Waters, K. A Prisoner's

Dilemma Experiment on Cooperation With People and Human-Like Computers. *Journal of Personality and Social Psychology 1996*. Vol. 70, No 1, 47-65

9. Kollock, P. Social Dilemmas: The Anatomy of Cooperation. *Annual Review of Sociology 1998*. 24:183-214

10. Kollock, P. 'An Eye for an Eye Leaves Everyone Blind': Cooperation and Accounting Systems. *American Sociological Review* 58(6):768-786. 1993.

11. Kollock, P. Cooperation in an Uncertain World: An Experimental Study. *Sociological Theory and Methods* 8(1):3-18. 1993.

12. Ledyard, J.Public Goods: A Survey of Experimental Research. In Kagel, J. and Roth, A. (Eds.), *The Handbook of Experimental Economics*. Princeton University Press, Princeton, NJ, 1995

13. Macy, M., Kollock, P. and Yamagishi, T. The 'Tiger's Cave' Experiments, *Cornell University Working Paper*, 1997

14. Messick, D and Brewer, M. Solving Social Dilemmas. In Wheeler, L. and Shaver, P. (Eds.), *Review of Personality and Social Psychology*. Sage, Beverly Hills, CA, 1983

15. Olson, M. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press, Cambridge, MA, 1965

16. Rocco, E. Trust Breaks Down in Electronic Contexts but Can Be Repaired by Some Initial Face-to-Face Contact, in *Proceedings of CHI '98* (Los Angeles CA, April 1998), ACM Press, 496-502.

17. Sally, D. Conversation and Cooperation in Social Dilemmas: A Meta-analysis of Experiments from 1958 to 1992. *Rationality and Society*, 7,58-92, 1995

18. Vellon, M., Marple. K., Mitchell, D., and Drucker, S. The Architecture of a Distributed Virtual Worlds System. *Proceedings of the 4th Conference on Object-Oriented Technologies and Systems (COOTS*). April, 1998.

19. Wellman, B. and Gulia, M. Net Surfers Don't Ride Alone: Virtual Communities as Communities. In Kollock, P. and Smith, M. (Eds.), *Communities in Cyberspace*. University of California Press, Berkeley, CA, 1998

20. Whittle, D. *Cyberspace: The Human Dimension*. W. H. Freeman and Co., New York, 1997

21. Wyer, R. S. "Prediction of Behavior in Two-Person Games." *Journal of Personality and Social Psychology* 13:222-238. 1969.

22. Yamagishi, Toshio. "The Provision of a Sanctioning System as a Public Good." *Journal of Personality and Social Psychology* 3:51-87. 1986.