Getting the Seats of Your Pants Dirty:

Strategies for Ethnographic Research on Virtual Communities

Luciano Paccagnella
Department of Sociology
University of Milan

Table of Contents

- Abstract
- Introduction
- Enduring Issues in CMC Research
  - Is Quality vs. Quantity a Real Dilemma?
  - Off-line and On-line Worlds
  - Naturalistic Analysis
  - Ethics of Research
- Ethnographic Research on Virtual Communities
  - Can a Computer Program Really Help?
  - Take, for Example, the Italian Cyber_punk Computer Conference...
  - A Proposal for Comparative Analysis on Virtual Communities
- New Perspectives: Toward the Study of Cultural Cyber-artifacts
- References
- Footnotes
- About the Author

Abstract

The study of social worlds built by people on computer networks challenges the classical dimensions of sociological research. CMC scholars are prompted to exploit the possibilities offered by new, powerful, and flexible analytic tools for inexpensively collecting, organizing, and exploring digital data. Such tools could be used within a Weberian perspective, to aid in systematic examination of logs and messages taken from the actual life of a virtual community. A proposal can then be made for a longitudinal strategy of research which systematically compares specific aspects of virtual communities over different periods of time and different socio-geographical contexts. The article summarizes a case study on an Italian computer conference, and concludes with a short outline of the new graphical CMC environments and their consequences for
the rise of a multimedia cyber-anthropology.

Introduction

Computer-mediated communication systems exhibit a fair amount of interpretative flexibility. That is, they can mean different things to different individuals or different groups, and their use continues to be interpreted and reinterpreted with the passing of time [Croft, Lea, & Giordano, 1994]. It is well known that the Internet was originally conceived as a military project supervised by the Defense Advanced Research Project Agency, created during the Cold War as an information system capable of surviving a Soviet nuclear attack [Miller, 1996]. Those same features of decentralization and flexibility that should make it militarily invulnerable contributed to giving us the Internet of today: an international, chaotic, dense bazaar inhabited by all kinds of people.

The interpretative flexibility of computer networks also means that academic researchers have been studying them from a wide range of perspectives and with a variety of methods: to name but a few, there have been ethnographic accounts of specific virtual places ([Baym, 1992]; [Myers, 1987]; [Reid, 1995]); evaluative studies of a system's costs and benefits ([Kerr & Hiltz, 1982]); analysis of intraorganizational networks ([Danowski & Edison-Swift, 1985]; [Rice, 1982]; [Sproull & Kiesler, 1986]); laboratory experiments comparing face-to-face and electronic communication ([Dubrovsky, Kiesler, & Sethna, 1991]; [Kiesler, Siegel, & McGuire, 1984]); hermeneutic interpretations ([Lee, 1994]); electronic surveys [Parks & Floyd, 1996], legal and normative analyses ([Maltz, 1996]; [Mnookin, 1996]), innovative gender studies ([Danet, 1996]; [Jaffe, Lee, Huang, & Oshagan, 1995]; [Matheson, 1992]) and so on. While of course a precise categorization of research in CMC would be impossible, it can be observed that scholars have lately shifted from the early studies of business communication to approaches which are more attentive to the symbolic and cultural dimensions ([Jones, 1995a]; [Lea, 1992]; [Mantovani, 1996]; [Sudweeks, McLaughlin, & Rafaeli, 1997]).

The very concept of interpretative flexibility is a cornerstone of a perspective based on social constructivism, which is not solely an approach to technology study ([Croft, Lea, & Giordano, 1994]; [Bijker, Hughes, & Pinch, 1987]), but also a wider epistemology conceived in reaction to the logical empiricist methodology and the bid to apply that framework to the social sciences [Schwandt, 1994]. Rooted in the phenomenological tradition of Alfred Schutz, in the philosophical pragmatism of George Herbert Mead and in the formal sociology of George Simmel, social constructivism claims that the facts of the world are not independent of us as observers and that scientific knowledge is always the result of a situated perspective. People create their own reality through an iterative process where man is at the same time producer and product of the social [Berger & Luckmann, 1966].

Anyone familiar with the pattern of life on Internet Relay Chat ([Bechar-Israeli, 1995]; [Reid, 1991]), on MUDs and MOOs ([Bruckman, 1992]; [Marvin, 1995]; [Mnookin, 1996]; [Reid, 1995]), on Usenet newsgroups ([Baym, 1992]; [McLaughlin, Osborne, & Smith, 1995]) or on BBSs [Myers, 1987] cannot fail to note that cyberspace constitutes a wonderful example of how people can build personal relationships and social norms that are absolutely real and meaningful even in the absence of physical, touchable matter. Theories of reduced social cues in CMC ([Dubrovsky, Kiesler & Sethna, 1991]; [Sproull & Kiesler, 1986]) have thus been questioned by a number of more constructivist-oriented studies ([Baym, 1995a]; [Lea & Spears, 1995]; [Mantovani, 1996].
1994]; [Myers, 1987]; [Spears & Lea, 1992]; [Walther, 1992]) whose common claim is that the availability of means of communication limited by characters typed on a computer keyboard does not prevent users from constructing their own social worlds. Social worlds, even in cyberspace, are exceedingly complex and their basic characteristics cannot be determined by any intrinsic feature of the communication medium: relationships on the net can be altogether more or less democratic, uninhibited or egalitarian than in real life, depending on an intricate pattern of elements. In fact, proponents of the SIDE model [Social Identity De-individuation (see Spears and Lea, [1992], [1994]) showed that in particular conditions on-line behavior can be even more social and normative than face-to-face interaction.

This article is intended to provide some methodological suggestions for the study of social worlds built by people on computer networks. Virtual communities has lately become a fashionable term which will be used here as a useful metaphor to indicate the articulated pattern of relationships, roles, norms, institutions, and languages developed on-line. This is not to say that we take the term virtual community as a positive value in itself, nor that we advocate an enthusiastic or optimistic view of computer networks. Even the very authenticity of communities developed on-line should not be taken for granted without an effort to come to a commonly accepted definition of what a community really is. The term virtual community is therefore still a problematic scientific concept ([Jones, 1995b]; [McLaughlin et al. 1995]). Anyway, communities are indeed worth studying when we do not look at them with romantic eyes, but with the eyes of the interpretivist ethnographer: according to Geertz [1973], man is an animal suspended in webs of significance he himself has spun and the job of the researcher is to achieve a thick description of those webs.

Doing ethnographic research on virtual communities requires different tools from business-organizational evaluation studies, for which excellent methodological literature already exists (see [Kerr and Hiltz, 1982]; [Rice, 1989]): emphasizing the community aspect (if any) of a computer conference, for example, might mean focusing attention on the subjects playing an active role as senders or recipients (and often both) of the messages. This clarification is important because other points of view could exist. The total readership of a computer conference is always somewhat larger than that of the actual active users; computer conferences are for some people just unidirectional information sources. Only a few persons (and it would be interesting to identify who) come to appreciate the conference as a social environment, where they acquire friends and enemies and build their own unique on-line identities.

While the goals of many early studies on CMC were related to the impact of new communication technologies in the efficiency of office work, a constructivist interpretation of virtual communities could come closer to the study of everyday life. Through the understanding of on-line social interaction we can also hope to be able to understand better the complexity of our daily social experience.

Enduring Issues in CMC Research

Is quality vs. quantity a real dilemma?

Despite the recent advances in the methods used in social sciences and the sophistication of post-modern epistemological debates, one of the first things most people still want to know when one speaks about social research is whether one’s orientation is quantitative or qualitative. The quality/quantity dichotomy has ruled the speculation on human inquiry throughout this century, creating opposing methodological schools and Weltanschauungen; the same dichotomy could now appear even in research on CMC.
What will be suggested in this article is, instead, that CMC constitutes a field which, given its own intrinsic characteristics, could transcend the traditional quality/quantity distinction, fostering at the same time new perspectives of analysis.

It is worth pointing out, by the way, that many of the shortcomings imputed to quantitative research as a whole are actually ascribed to particular practices: for example, the concern that when the researcher interviews people following a structured questionnaire he or she usually lacks the ethnographic context necessary to understand the answers [Cicourel, 1974]; or, a common feeling that quantitative researchers assume that analyzing a great number of cases from a large-scale administration of a survey would automatically reveal something useful about the phenomena being studied, even though the sampling process might be biased [Schwartz & Jacobs, 1979].

Put in other words, the quality/quantity dichotomy often obscures legitimate concerns that researchers are drawn to designs that oversimplify social reality and take little notice of the sense and meaning of situations from the standpoint of the actor. Still, the idea of an interpretive social science that, where appropriate and with the necessary caution, also makes use of statistical methods, is not a novel one: even Max Weber (the father of systematic, interpretive methods in social sciences) advocated an epistemological approach where quantitative measurements were not excluded a-priori and where scientific explanation (Erklären) and interpretive understanding (Verstehen) could support each other [Weber, 1922]. This classical German scholar would probably remind us that there's no need to rail at quantity when we actually aim our criticisms at the positivist illusion that some intimate and definitive essence of reality can be captured. Qualitative and quantitative methods have often been associated with mutually exclusive views about society and social science, and there is no doubt that qualitative methods have encouraged a non-positivistic practice of science. While we do believe that mutually exclusive views still exist, we are not sure about the correctness of the simple distinction between quality and quantity. Perhaps wisdom lies in being tolerant and shamelessly eclectic in our use of methods [Rossman & Wilson, 1994].

Long-lasting Weber's thought can teach us even a further lesson: just as mechanical, heavy quantification may only lead the researcher far away from the phenomena he/she wants to study, in the same way a naive application of the concept of Verstehen must be avoided: to understand the meaning of an event is not a process of empathic identification or a psychological getting inside someone's head. Weber rejected the romantic version of Verstehen as conceived by eighteen-century German philosophers (in particular by Wilhelm Dilthey), which was close to the concept of fühlen (that is to feel through emotions) and which implied a comprehension totally based on immediate and intuitive mechanisms. Weber's understanding is far less romantic, grounded as it is on a conceptual re-construction of meaning achieved little by little by the systematic analysis of texts and documents.

Deep, interpretive research on virtual communities could consequently be greatly helped by an accurate use of new analytic, powerful yet flexible tools, exploiting the possibility of cheaply collecting, organizing and exploring digital data. Some suggestion of what these tools could look like will be given later on in this article.

**Off-line and on-line worlds**

While debate over quantitative and qualitative methods may continue as a thread in discussions of research on CMC, new kinds of problems remain in the background. There has been little discussion of the assumption, for example, that an electronic survey among the users of a given system could provide a cheap and fast way to draw
connections between on-line behavior and traditional socio-demographic variables (age, gender, level of education, family income, etc.). But what is the real meaning of socio-demographic data obtained through, say, a structured on-line questionnaire? What is really happening, for example, when SweetBabe, a regular participant in IRC channel #netsex and one of the hypothetical cases from our survey sample, tells us that her real name is Mary, she's thirty years old and she works as a secretary?

It is wise to suppose that, more than providing us some (if any) actual information about Mary's real life, such an answer could help to understand better SweetBabe's symbolic universe, her on-line self-representation, her social values and relationships. In a perspective of ethnographic research on virtual communities the on-line world has its own dignity: after all, from a phenomenological standpoint, SweetBabe and her social world are for us much more real than this supposed Mary about whom we actually know absolutely nothing. Even when the design of research does expect some data referring to the real world, it is never correct to accept these data without keeping in mind that obtaining information about someone's off-line life through on-line means of communication - although seemingly easy and convenient - is always a hazardous, uncertain procedure, not simply because of the risk of being deliberately deceived but also because in such cases the medium itself increases the lack of ethnographic context discussed above and it may also produce misunderstandings due to different communication codes.

Finally, it could be useful to note that many of the most interesting virtual communities are also very proud of their exclusive culture. A stranger wanting to do academic research is sometime seen as an unwelcome arbitrary intrusion. Things then become similar to researching in difficult environments such as extremist political groups, and specific precautions have to be taken [Diani & Eyerman, 1989].

Another example where the complexity and richness of on-line social worlds have often been underestimated is in the analysis of power and status relationships. Well-known laboratory experiments comparing face-to-face communication with electronic mail found that computer networks have a status equalization effect [Dubrovsky, Kiesler & Sethna, 1991]; a few field studies confirmed that organizational electronic mail reduces social differences and increases communication across social boundaries (e.g. [Sproull & Kielser, 1986]). The technological determinism hidden behind these positions has nonetheless been questioned by other scholars (e.g. [Baym, 1995a]; Mantovani, [1994], [1996]; [Myers, 1987]; Spears & Lea, [1992], [1994]), highlighting the importance of the social context surrounding the use of the medium.

Furthermore, moving away from strictly organizational task-related experiments, more social-oriented ethnographic studies on CMC have appropriately identified the existence of strategies of visibility of the actors which make up for the lack of traditional interpersonal cues and which indeed permit the development of a status differentiation ([Bruckman, 1992]; [Meyer & Thomas, 1990]; [Myers, 1987]; [Reid, 1991]) of the newcomers to a computer conference or a MOO are immediately recognized as such and the same holds true for the leaders. Both acquire and use symbols that make them different one from the other even if they are all apparently hidden beyond the keyboard of one's own computer. Such a status differentiation, of course, may not match a pre-existing differentiation in the off-line life, if any; this suggests that maybe theoreticians of equalization effects are more interested in verifying the persistence of traditional power relationships in CMC rather than in recognizing the spontaneous increase of new specific norms, identities and relationships.

Naturalistic analysis
The availability and accessibility of unobtrusive techniques constitutes a further reason to question the simple application of strategies like the survey to the study of virtual communities. It is well known how, in social sciences as well as in other fields, the phenomena being studied are modified by the very act of observing them. Even in the case of soft, qualitative techniques, as in participant observation, problems arise because of the presence of the researcher in the field. Kerr & Hiltz [1982], for example, reported two problems specifically related to participant observation in CMC: going native and role conflict, the first referring to involving oneself in the group to the extent that objectivity is lost, while the second means a dilemma between the goals of the group and those of the evaluation.

Although doing a good job of systematic observation is without doubt much more difficult and problematic than simply walking around and describing what one sees, it is still true that in the study of virtual communities we can exploit some peculiar advantage. In many cases observation can be carried out even without informing the people being studied. While this obviously urges us to take into consideration new ethical issues (that will be discussed further on), at the same time it reduces the dangers of distorting data and behavior by the presence of the researcher. Despite the fact that observation in nature is usually also economically much more convenient compared to the usual budgets of empirical research, often the tradition (in particular of economists and social psychologists) leads us to prefer laboratory experiments carried out with small groups of people recruited for the occasion, typically university students (e.g. [Dubrovsky, Kiesler & Sethna, 1991]; [Kiesler, Siegel & McGuire, 1984]).

But laboratory experiments have been conducted primarily in task-oriented research on CMC, and their findings cannot be generalized to all domains of electronic communication. The shortcomings of such experiments have been pointed out by scholars interested in studying virtual communities: for example, Nancy Baym [1995a] highlights that differences in experimental research designs (i.e. characteristics of the group and the members, tasks required of the group, communication systems, and groups' temporal structure) are rarely addressed, oversimplifying the forces that affect CMC; she then calls for more flexible and dynamic approaches, such as her own naturalistic study on a specific newsgroup (Baym, [1992], [1995b]). David Myers tried to set up a game experiment among regular users of a Bulletin Board System (BBS), finding that it was not useful in discovering the basic process by which communication strategies evolve: as the artificial and predetermined nature of experiments does not permit the users to manipulate the context creatively - then it's not surprising if the social-related aspects of CMC have hardly emerged in laboratories; mechanisms of context manipulation have instead been wonderfully explained by Myers's participant observation [Myers, 1987]. Spears and Lea, reviewing the equalization effect and the reduced social cues approach (see e.g. [Dubrovsky et al., 1991]; [Kiesler, Siegel & McGuire, 1984]; [Sproull and Kiesler, 1986]), argue that the unfamiliarity of many participants with CMC in experiments may have lead to biased conclusions, and suppose that findings of laboratory experiments might not be confirmed in the field anyway, where social relations are long-term and have long-term consequences [Spears & Lea, 1994]. Therefore the invitation issued by Robert Park in the first half of this century to get the seat of our pants dirty by real research conducted out of the classrooms and laboratories is still valid.

Structure of time seems to be a particular source of flaws in experimental studies. A meta-analysis of previous research [Walther, Anderson, & Park, 1994] argued that time constriction implicit in controlled laboratory conditions could be held responsible for the contradictory characterizations of CMC in controlled experiments as opposed to field studies. Walther and his collaborators calculated that the median time allowed for CMC
in time-restricted experiments was 30 minutes. This induces two considerations: a) since communication in CMC has been reported as being usually slower than in face-to-face interaction ([Lea, O'Shea, Fung & Spears, 1992]; [Walther, 1992]), experiments allowing restricted equal time periods in CMC and in face-to-face groups do not take into account the possible interaction of time limits with the communication channel difference; b) the typical setting of a laboratory, where people are asked to perform a given task in a given time, mostly without pre-experimental relationship history and without expectation of future interaction, can itself hardly be considered a good environment to observe the social richness and complexity of an established group.

Reflection about temporal structure in CMC should also prompt a consideration of the general limits of a static research which will not take the trouble to discover and understand the dynamic and the slow evolution of the specific culture and social climate of a particular piece of cyberspace. There is therefore need for future research that, besides being conducted in the field and not in laboratory conditions, is also longitudinal.

**Ethics of research**

Field research conducted with unobtrusive techniques is inevitably doomed to create major ethical problems. Scholars generally do not agree on common ethical guidelines: some feel that they have a moral obligation to obtain explicit permission from the authors for publishing logs in academic papers (e.g. [Marvin, 1995]); others collect logs without asking for permission but the logs are then only processed by statistical software and not read by humans [Danowski & Edison-Swift, 1985]; many others simply do not declare explicitly whether permission was obtained for their logs (e.g. [Reid, 1991]). All, though, are concerned with the privacy of the users and do take precautions such as changing names, pseudonyms, or addresses from the logs. Changing not only real names, but also aliases or pseudonyms (where used) proves the respect of the researchers for the social reality of cyberspace.

The Forum on The Ethics of Fair Practices for Collecting Social Science Data in Cyberspace, recently organized by Jim Thomas as a special section of the scholarly journal The Information Society, illustrates the variety of positions about ethical guidelines in on-line social research, identifying a deontological, a teleological and a postmodern approach [Thomas, 1996].

Ethical concerns have also been discussed at length in the ProjectH Research Group, a team of scholars from several countries and many universities that collaborated in 1993-94 on a quantitative study of electronic discussions. After lengthy debate the group voted an ethical policy that would not seek permission for recording and analysis of publicly posted messages:

*We view public discourse on CMC as just that: public. Analysis of such content, where individuals', institutions' and lists' identities are shielded, is not subject to 'Human Subject' restraints. Such study is more akin to the study of tombstone epitaphs, graffiti, or letters to the editor. Personal? - yes. Private? - no* [Sheizaf Rafaeli, as quoted in Sudweeks & Rafaeli, 1995].

Although the study aimed to treat messages quantitatively and its policy cannot therefore be accepted as a universal ethic, it seems that the ProjectH group took a reasonable position. Conversation on publicly accessible IRC channels or messages posted on newsgroups are not equivalent to private letters (while private, one-to-one e-mail messages of course are); they are instead public acts deliberately intended for public
consumption. This doesn't mean that they can be used without restrictions, but simply that it shouldn't be necessary to take any more precautions than those usually adopted in the study of everyday life. As Jim Thomas clearly explained:

_Eavesdropping on a private conversation, even in a public place, is surreptitious, and therefore unethical. Tape recording a private conversation unbeknownst to the participants is illegal. (...) However, if I set up video camera to record activities in the park, and if the seated friends are part of the panorama, and if the friends suddenly began shouting loud enough to be recorded from a distance, *that* would be fair game._

[Message posted to the cybeth-l mailing list on Dec. 14th, 1995]

Ethnographic Research on Virtual Communities

Can a computer program really help?

Although surveys and quantitative sociology have long been carried out with the help of specific software, interpretive research has often seen computers as tools too rigid to have anything to do with the complexity of social phenomena.

Let's now imagine a computer program that would permit:

1. Making notes in the field
2. Writing up or transcribing field notes
3. Editing: correcting, extending, or revising field notes
4. Coding: attaching keywords or tags to segments of text to permit later retrieval
5. Storage: keeping text in an organized database
6. Search and retrieval: locating relevant segments of text and making them available for inspection
7. Data linking: connecting relevant data segments to each other, forming categories, clusters, or networks of information
8. Memoing: writing reflective commentaries on some aspect of the data as a basis for deeper analysis
9. Content analysis: counting frequencies, sequences, or locations of words and phrases
10. Data display: placing selected or reduced data in a condensed, organized format, such as a matrix or network, for inspection
11. Conclusion-drawing and verification: aiding the analyst to interpret displayed data and to test or confirm findings
12. Theory-building: developing systematic, conceptually coherent explanations of findings; testing hypotheses
13. Graphic mapping: creating diagrams that depict finding or theories
14. Preparing interim and final reports

These are, in fact, the main uses of computer software in qualitative studies nowadays as outlined by Miles and Huberman [1994]. Add a few other features such as: collecting and archiving data in automatic (or semi-automatic) and unobtrusive ways; keeping any available information (e.g. the sender, recipient, subject, date and time of an electronic message) in different logical fields, in order to allow sophisticated and precise Boolean searches; accepting several types of wildcards in searches and displaying the results in a user-definable context; exporting frequencies and other quantitative data in format(s) suitable for major statistical packages.

A computer program implementing all these features would indeed be of some help. As Weitzman and Miles [1995] note, no computer program will _analyze_ your data:
Computers don’t analyze data, people do. Still, just like guns do make it easy for people to kill other people, so a well-designed computer program could make it much easier for the researcher to think about the meaning of his/her data. Unfortunately, to the best of our knowledge, a unique software implementing all these features does not exist yet. However, there are several products that are close to our ideal program, each of them performing some of the tasks described above. Most of these programs are labeled under the growing category of software for qualitative analysis that is gaining more and more attention since the diffusion of personal computers led to a phenomenal development in this domain (e.g. [Richards & Richards, 1994]; [Tesch, 1990]; [Weitzman & Miles, 1995]).

Although this kind of software usually provides many of the features that we need, the expression software for qualitative analysis sounds inadequate: our aim is, more generally, to acquire information effortlessly with the aid of a computer program, to study and interpret this information with the aid of our own hermeneutic understanding directly earned in the field, and to produce some clear synthesis of the results. We do not see any need to decide, a-priori and once for all, whether data analysis and presentation of results will or will not also make use of numbers. On the contrary, an ideal software for the analysis of virtual communities would offer the possibility of switching between numerical and textual/contextual descriptions of data with little effort, freeing the researcher from technical constraints and letting him/her have the freedom to try out different strategies.

Take, for example, the Italian cyber_punk computer conference...

We can now provide a practical example of how socio-anthropological research can be conducted following the guidelines discussed above. The example is a study of an Italian virtual community built around a computer conference named cyber_punk. This conference, hosted by a non-profit BBS network, was chosen as a case study for several reasons: first of all, it is an experience which began and was developed in Italy (although there was considerable contact and collaboration with similar experiences abroad) and this should help to eliminate (partially at least) the continuous reference to the American scene; in contrast with the rampant fashion of the Internet, this is a poor enterprise, because it uses cheap technical tools accessible to everyone; it tries to address a popular communication demand, it is open to everybody who owns a computer, a modem and a telephone line, and it does not require access fees other than those imposed by the telecom company for the use of the telephone line; it is a conference without a specific topic and a very precise goal, but it has nonetheless been able to develop a real identity and a very strong community feeling; finally, the Cybernet network which hosts the conference presents some features which are quite interesting to the social scientist: for example, there is not any network policy, just as there are not any specific requirements to be met for the admittance of new nodes. The hope is to stress the idea of the network as a space, cyberspace indeed, where everyone is responsible for his or her own actions and for maintaining his or her own image as effectively as possible in that specific social context ([Goffman, 1959]; [Hiemstra, 1982]). Each node finds its own internal management policies, is self-financing (the system operator can pay for everything, or can ask users for help) and chooses which internal rules to observe.

In order to gain an intimate understanding of the culture and the symbolic system of the conference, the author has been a participant observer for 18 months. All the messages have been recorded and archived every month in a separate file, for a total of nearly 10,000 messages and 400 users involved. Permission to record the data was not sought, although many users were aware of the study and the study itself has never been hidden or concealed, its nature being declared to the participants concerned on request.
ASCII data files have been subsequently imported by a textbase manager, specifically customized to read the raw files and to automatically import the information into appropriate fields without manual intervention [1].

The textbase manager proved to be a fundamental tool indispensable to add, to the confidence acquired through direct participation, a precise and objective reference to the collected data. The risk of going native [Kerr & Hiltz, 1982] in participant observation is related to the loss of a neutral perspective in describing and analyzing data. In this case the researcher exploited the possibility of an extremely simple, fast, and effective exploration of the thousands of collected messages: multiple windows showed on the screen the results of searches for particular situations, for example the dialogues between specific groups of actors, in a given period, on a particular topic.

A few small computer programs, specifically developed for this task, converted the raw ASCII data files into a format suitable for a suite of statistical packages: a simple content analysis program [2], a program for communication network analysis [3] and the well known SPSS. Although the results of statistical analyses have been reported extensively elsewhere [Paccagnella, 1996], it is worth noting that every effort has been made to keep the researcher near the data and not to let numbers dominate the scene.

This integrated approach permitted the acquisition of some confidence in the basic social characteristics of this virtual community, including its peculiar slang, the role structure, socialization processes and the evolution over time of the overall social climate of the conference. It has been possible to understand and describe the expectations related, for example, to the role of leader, whose linguistic competency is far from being simply that of a computer nerd and refers instead to a collective sense of identity. Analogical communication has been studied both in excerpts from actual messages and in its statistical correlation with other measurements of social liveliness, revealing the importance of the use of smileys, interjections, and informal styles of conversation in the building of a strong and culturally vivacious community.

However, while a better theory of how virtual communities arise and develop has still to be reached, more coherent techniques of research and software tools are required as well.

A proposal for comparative analysis on virtual communities

Emphasizing the role of specific software in the management of automatically recorded messages and logs brings with it the danger of turning back to a positivistic view of research: it is not safe to think of these data as some sort of objective reality frozen by the computer. Archived messages and logs are representations of the on-line phenomena as perceived by participants. The actual social reality lying behind these phenomena is more complex in that logs lack at least two aspects of interaction [Marvin, 1995]: first they do not record the dynamic dimension of turn-taking, which can occur over a few seconds (in synchronous CMC, i.e. on IRC) or over several days (in asynchronous conferences and mailing lists) - and sometimes the typing time or the lags play a fundamental role in shaping the collective mood surrounding the messages. Secondly, logs ignore the actual experiences of individual participants at their own keyboards in their own rooms all around the globe. Reid [1995] suggests another consideration about the shortcomings of logs taken from MUD interaction, which can be applied to asynchronous CMC as well: the language of computer-mediated communication is more ephemeral than ordinary written texts (it has therefore been defined as written speech); it is not intended for people uninvolved directly in interaction, and it loses part of its sense and meaning when re-read afterward by neutral observers.
For this reason the research design suggested in this article is theoretically rooted in Weber's notion of *Verstehen* (and in its fertile subsequent re-conceptualization throughout this century in sociology and anthropology) and it exploits the possibilities of analytical, systematic tools for collecting and analyzing machine-readable data without trusting them as objective or self-revealing slices of reality. The attention is primarily focused on a single case: a newsgroup, a mailing list, a BBS conference, an IRC channel, a MUD, or any other virtual community. To understand the intimate social processes, the cultural and symbolic systems of any group requires much time and effort, and CMC scholars have by now provided us with a fair number of lovely, accurate descriptions of specific venues in cyberspace (e.g. [Baym, 1992]; [Bruckman, 1992]; [Meyer & Thomas, 1990]; [Reid, 1991]).

We could now move one step further, beginning to think about a comparative, longitudinal strategy of research on virtual communities. Comparative methods have been discussed since the rise of social sciences, sometime making them the main way to scientific explanation. In fact, describing a situation (i.e. saying that it is *democratic* or *task-oriented*) means comparing it with others [Smelser, 1976]. Two of the fathers of sociology, Emile Durkheim and Max Weber, have also been masters of comparative studies.

Comparative analysis of virtual communities could be close to a case-based strategy, mostly associated with Weber's perspective [Ragin & Zaret, 1983] and therefore consistent with our overall interpretive approach. Comparing a few virtual communities, carefully chosen in different geographical, cultural, and technical contexts, could prove a good method to obtain some first generalizations on human social behavior in cyberspace. Hypotheses suggested by deeply conducted case studies could be tested and verified in completely different conditions, allowing the researcher to control and rule out variables mostly related to the outside off-line world. It would become possible to draw general considerations about the basic processes of cooperation in natural groups on computer networks, including: the development and use of local slang; the socialization and acceptance of newcomers; the stigmatization of the *outsiders*; the rise of an élite of leaders, its legitimation and renewal; the approval of a written or unwritten corpus of social norms; the responses to disruptive behavior; the persistence of relationships developed on-line in other venues of cyberspace or even in the off-line world, and so on.

An international, informal team of researchers, each of them bringing his/her own understanding and experience about a specific community, could draw a comparative, longitudinal project of research on CMC. Data (consisting mainly of messages and logs but also multimedia files such as images, sounds or animations) can be systematically collected, organized and shared with the aid of computer programs specifically customized for this task. This kind of collaborative comparative design would permit taking into consideration the social contexts of each case, working, by the way, not on a sample of messages or logs but on the whole population.

**New Perspectives: Toward the Study of Cultural Cyber-artifacts**

Ethnographic studies of CMC have so far been studies of *text-based virtual realities* ([Bruckman, 1992]; [Marvin, 1995]). According to Reid [1995] virtual reality is primarily an imaginative rather than a sensory experience and the paraphernalia commonly associated with immersive virtual reality (data gloves, head-mounted displays, etc.) do not account for the acceptance of a simulated world as a valid site for emotional involvement. Text-based worlds are not surrogates for audiovisual experiences and it seems reasonable to suppose that in certain situations people will still...
prefer textual interactions even when more advanced technologies become widely available in the future.

Technologies like CU-SeeMe or Internet Phone are already allowing people to see and hear each other through the Internet but, apart from their poor quality compared to the analog video and phone systems, they lack some of the peculiar features of text; text guarantees everyone the possibility to be heard, while multimedia technologies are suitable only for small groups [Bechar-Israeli, 1995]: after all, even in face-to-face interaction you can listen to no more than two or three people speaking at a time. Shakespeare and the other classics of literature can teach us how text is able to express emotions, experiences and complex ideas, and the fact that filmmakers sometime still choose to shoot movies in black and white demonstrates that narrowing the bandwidth often helps in focusing the message [Godwin, 1994].

Nonetheless, one cannot fail to note that the world of on-line communication is moving toward multimedia systems. Just as early text-based arcade videogames have been replaced by 3D graphic adventures, several types of graphic MUDs - where textual descriptions of personae and places are replaced by their graphical representations - are now preparing to capture the attention of the general public [Rossney, 1996].

Research on virtual communities cannot ignore these new environments, which can potentially take the task of the researcher extraordinarily close to that of traditional field anthropologists. Screenshots taken from graphical worlds like Alphaworld [4] show a land inhabited by different people and modeled by different artifacts: buildings, streets, gardens, means of transportation and other tools may be analyzed in their shapes and aesthetics. They can be captured and compared by specific software, perhaps like that discussed above for textual CMC.

Research on virtual communities will then be even more similar to research on traditional communities in "real life." Perhaps advanced concepts and theories specifically developed on CMC will find some application in mainstream social sciences, partially repaying the debt of computer-mediated communication research to older and more eminent disciplines.

References


• **Mantovani, G. (1994).** Is computer-mediated communication intrinsically apt to enhance democracy in organizations?. *Human Relations, 47, 1*, 45-62.


• **Miles, M. & Huberman, A. (1994).** *Qualitative data analysis: An expanded..."


- **Myers, D. (1987).** Anonymity is part of the magic: Individual manipulation of computer-mediated communication context. *Qualitative Sociology, 10*, 251-266.


Footnotes

[1] The software used was Folio VIEWS, version 3.01 for Windows, developed by Folio Corporation.

[2] Textpack V, version 3.0 for personal computer; this program was developed by a research team at the Zentrum für Umfragen, Methoden und Analysen in Mannheim, Germany.

[4] Alphaworld is just one of the several graphic multi-user environments available on the net. More information at http://www.worlds.net/alphaworld/

About the author

Luciano Paccagnella is a doctoral candidate in Sociology at the University of Milan (Italy). He graduated in 1994 at the University of Trento (Italy) with a dissertation titled Sociology of Cyberspace: The Social Construction of Reality on Computer Networks. He has been involved as a system operator in grassroots computer networks and is currently member of a team investigating cases of civic networking in Italy and abroad. His research interests include new social movements and innovative methods of social research. Mr. Paccagnella is also writing his doctoral dissertation on virtual communities.