

Research, Methodology and the Internet:

A study of the Internet as a data capturing tool

Zoë Teresa Dowling

Thesis presented in partial fulfilment of the requirements for the
degree of Master of Philosophy at the University of Stellenbosch



Supervisor: Professor Johann Mouton

December 2001

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

ABSTRACT

It is widely accepted that the Internet has become a valuable resource for social scientists, not just for the purpose of information exchange; via e-mail, discussion groups and electronic journals, but also as a medium for data collection. Its global nature gives a researcher access to a vast range of individuals located around the world. It also opens up access to difficult to hitherto penetrate study areas, such as sensitive research on deviant behaviour. Further, it is claimed that considerable savings to both research budgets and time frames are made possible with the new technology.

It is not surprising, therefore, that a substantial body of research, employing the Internet as the primary means of data collection, already exists. This raises a number of questions as to how the Internet fares as a research tool. Are there any important methodological issues that this new approach raises? Do the traditional research methods suffice? Or are adaptations to existing methods necessary when difficulties are encountered? Does such changes affect the more fundamental question of the research design? These questions are considered in this thesis.

To answer them, I consider two different types of empirical research designs. The first, survey research, is a quantitative, numerical design that traditionally has a high level of control. I consider in detail issues of sampling, including non-response, and questionnaire design. The second design examined, ethnographic research, is qualitative, textual and generally has a low level of researcher control. I address the methods used in cyber ethnography and then discuss the considerable ethical concerns that feature in such research.

I conclude that, on the whole, the existing methods can be transferred to Internet research. Indeed, some of the problems faced in traditional research are also considerations in Internet studies and can be overcome by employing similar techniques, such as using incentives to reduce non-response rates. However, a number of new problems emerge, such as the lack of paralinguistic cues, which require adaptations to the existing methods in order to produce results that can be considered valid and reliable. However, I also argue that these necessary adaptations to the methods do not affect the underlying principles found in the research design.

OPSOMMING

Dit word algemeen aanvaar dat die Internet 'n waardevolle hulpmiddel is vir sosiaalwetenskaplikes; nie net vir die uitruil van inligting deur middel van e-pos, besprekingsforums en elektroniese joernale nie, maar ook as 'n instrument om inligting te versamel. Die globale aard van die internet gee die navorser toegang to 'n wye spektrum individue internasionaal. Dit verleen ook toegang tot moeilike navorsingsareas, soos sensitiewe navorsing oor afwykende gedrag. Verder word beweer dat dit aansienlike besparings moontlik kan maak vir beide die navorsingsbegroting en tydraamwerk.

Dit is dus nie verbasend dat 'n substansiële hoeveelheid van navorsing, wat die Internet as die primêre bron van dataversameling gebruik, reeds bestaan nie. Dit laat verskeie vrae ontstaan oor hoe die Internet vaar as 'n navorsingshulpmiddel. Is daar enige belangrike metodologiese kwessies wat hierdie nuwe metode aanraak? Is die tradisionele metodes voldoende? Of moet daar veranderinge aan die huidige metodes aangebring word wanneer probleme ontstaan? Sal hierdie veranderinge die fundamentele aspekte van navorsingsontwerp beïnvloed? Die vrae sal oorweeg word in hierdie tesis.

Ek gebruik twee empiriese navorsingsontwerpe om die vrae te beantwoord. Die eerste, steekproefnavorsing, is 'n kwantitatiewe ontwerp wat tradisioneel 'n hoë vlak van beheer toon. Ek ondersoek in detail kwessies van steekproewe, insluitend geen respons en vraelysontwerp. Die tweede ontwerp wat ondersoek word, etnografiese navorsing, is kwalitatief, tekstueel en toon in die algemeen 'n lae vlak van navorser

beheer. Ek ondersoek die metodes wat gebruik word in kuberetnografie en bespreek dan die etiese vraagstukke wat hierdie navorsing kenmerk.

Ek kom tot die gevolgtrekking dat oor die algemeen die huidige metodes toegepas kan word op Internetnavorsing. Inderdaad kan van die probleme wat ondervind word in tradisionele navorsing ook ondervind word in Internet studies en ook hier kan dit oorkom word deur die gebruik van soortgelyke tegnieke, soos om aansporingsbonusse om geen responskoerse te verminder, hoewel daar nuwe probleme opduik, soos die gebrek aan para-taalkundige wenke. Dit noodsaak veranderinge aan die huidige metodes om resultate te lewer wat geldig en betroubaar is. Ek redeneer egter ook dat hierdie nodige veranderinge aan die metodes nie onderliggende beginsels van navorsingsontwerp verander nie.

ACKNOWLEDGMENTS

I am greatly indebted to my supervisor, Johann Mouton, not just for the advice and comments leading to the completion of this thesis, but also for all that I learnt in the Centre. Next I would like to offer thanks to Marthie van Niekerk for administrative assistance and more importantly, much needed moral support during those long days! Finally I dedicate this thesis to my parents, Dolina and Keith Dowling, who have unfailingly helped me along the way.

TABLE OF CONTENTS

Introduction	page: 1
Chapter One:	
Background to study	page: 4
Chapter Two:	
A closer look at Survey Research on the Internet	page: 17
Chapter Three:	
Internet Ethnographic Research	page: 56
Chapter Four:	
Conclusion	page: 93
References	page: 102

INTRODUCTION

Undoubtedly, in the 1990s, the Internet brought a major change to the lifestyles of many people throughout the world. Prior to this, few people had heard of it, let alone spend hours of work and leisure time 'surfing' the network for information, news or entertainment. However, nowadays many people shop, keep abreast of the latest national and international news, monitor their bank accounts or stocks and shares and communicate with people worldwide, all by the click of a mouse. Indeed, it has been recently estimated¹ (Nua; November 2000), that there are currently 407.1 million people who regularly log-on to this global network. It cannot be denied that a great number of those people do live in first world countries. However, there are indications that growth in Internet connectivity is being experienced by in all countries worldwide, albeit slowly. The same company estimates that there are around 4 million Internet users in Africa and 16.5 million in Latin America. This is compared to over 110 million in Europe and nearly 170 million connected in Canada and USA, clearly illustrating the dominance of the west.

We have witnessed, particularly in the last five years, an impressive growth in the application of this new technology in the Social Sciences. Anthropologists, psychologists, social historians and sociologists are eager use the network, and for good reasons. Just to mention a few: its global nature brings a new dimension to social studies. Prior to the Internet, social research was tied to a geographically limited population, often to those in close proximity to the researcher. Now,

¹ Nua (an Internet survey firm) report this to be an 'educated guess' compiled by observing many of the published surveys over the previous two years.

hypothetically, a researcher can access people in almost every country of the world. Similarly, research in socially sensitive areas, such as data dealing with intimate domestic matters or discussion groups concerned with sensitive medical issues, which hitherto the researcher may have been denied access, are now more readily contactable thanks to the Internet. Lastly, the Internet is cost-effective. It costs a user no more to contact the opposite side of the globe than it does the other side of town. The financial benefits to the researcher here are obvious.

Before looking in detail at the range of other uses of the Internet and the new problems it poses for social scientists, I will outline the way in which I have structured this thesis. In Chapter One, I present a brief outline of the history of the Internet, illustrating how it developed and grew to become a globally acclaimed network. Secondly, I briefly highlight its importance to the academic world generally and then outline some of the opportunities this new technology seems to offer to the social scientist. This includes the Internet as an information resource, on the one hand, to the emergence of the Internet as a new and very powerful data collection tool on the other. Surveys and ethnographic investigations are undertaken using the new technology. A detailed examination of the methodological implications of these studies is the focus of this thesis.

In Chapter Two, examples of research, where data has been collected by means of Internet surveys, are examined. I discuss two types of problems that are regularly encountered in such research. Issues pertaining to sampling are exceptionally important in the successful collection of valid, reliable and generalisable survey data. A range of conventional sampling procedures, viz. probability and non-probability

sampling, is first considered, followed by an investigation as to how these techniques have been employed when the Internet is used for data-collection purposes. In this section, I also address the problem of non-response in Internet surveys. The second problem in this chapter concerns questionnaire design in Internet research. This includes the advantages and limitations of both e-mail and World Wide Web based questionnaires.

In Chapter Three, we consider online (cyber) ethnographic research. In the first section, I address the practical (i.e. technological) features and techniques in online observation and interviewing in the ethnographic method. I also address issues of researcher reflexivity and interpretation in cyber ethnography. The second section in this chapter deals with a few of the extensive ethical considerations of such research. This includes discussions on informed consent, public and private spheres in cyber space and benefit and harm in social research.

Finally, in Chapter Four I draw together the findings of each chapter to present a comprehensive overview of the methodological advantages and limitations of cyber-based research. I argue that the Internet is a powerful new tool for the social researcher and its effectiveness as a tool, moreover, will increase as the technology develops and the diversity, and numbers, of people connecting to the network increases.

Chapter One

BACKGROUND TO STUDY

Section one: History

The Internet had its origins in the mid-1960s with the introduction of a United States military project, DARPA (Defense Advanced Research Projects Agency), which aimed to develop a robust communications system that could withstand sizeable losses to the underlying network and remain functional (Leiner *et al*; 2000:14). This, amongst other objectives, entailed the development of a computer network allowing information from computers, in different geographical locations, to be transmitted. A core feature of this network meant that if the most direct route (from one computer to another) in the system was not available, an alternative route would be utilized and communication/information exchange would remain possible (Lynch; 1993:5).

The first connection of this computer network, and hence the beginning of the Internet as we know it today, linked together four universities' computer systems in the United States – UC Los Angeles, Stanford Research Institute, UC Santa Barbara and Utah University (Howe; 1999). This took place in December 1969. By 1972, more than 37 universities in the US were linked to this network, then known as the ARPANET, with the numbers rapidly increasing thereafter (Winder; 1994:2). Although this may have been the beginnings of the Internet, it bore very little resemblance to the system we know today. In these early days, the network was used only by experts – computer professionals, engineers and scientists, who all had to learn to manoeuvre themselves around a very complex, non-user friendly system (Howe; 1999).

In the twenty years following the linking of these university computer networks, the system experienced tremendous growth with various agencies devoting large amounts of time and funding to its development. However, at this stage the most rapid expansion was seen within the US. At the beginning of 1983, the TCP/IP (Transmission Control Protocol/Internet Protocol) architecture, developed by Kahn and Cerf at Stanford, was universally adopted. In 1985, the (American) National Science Foundation (NSF) created the NSFNET, which consisted of five supercomputer centres containing resources that were available to educational institutions. The NSFNET later took over the ARPANET, whose defense background meant that its information was surrounded in "red tape" (Winder; 1994:3). In 1986, the NSF provided funding for a cross-country 56 kbps backbone for the Internet (Howe; 1999). This backbone allowed for the development of the Internet, as it soon became officially known. The financial support, provided by the NSF for the next ten years (totalling \$200 million), further facilitated its rapid growth (Howe; 1999).

During these years, several European nations developed their own internal networks – each with their own aims. Often these networks were created for specific communities. A prime example of this is JANET (Joint Academic Network) in the United Kingdom, which provided an inter-connecting wide-area network for the academic community (Leiner; 1993:18). Commercial networks were also developed at this time; for example AT&T's USENET and IBM's SNA (Leiner *et al*; 2000:8).

As it developed, the Internet became more user friendly. Many of the commands for various programmes became standardised (e.g. for e-mail, telnet and FTP) thus increasing the accessibility of the network to many more people. Nevertheless, even

in the early 1990's, the system was not as easily navigated as it is today (Howe; 1999). Many different organisations and universities worked at advancing this new technology in order to make it not just user friendly but also to develop the best way to utilise the tremendous potential that the network promised. Various attempts to create indexes for the information available on the network, saw the formation of applications such as Archie (McGill University) and Gopher, which were systems originally developed to aid access to files and information on the University of Minnesota's local network. Gopher soon grew into an application that allowed the user to move through many different networks (not just Minnesota University's network) in the search for information from one basic menu (Winder; 1994:5). Further enhancements of Gopher were created with VERONICA (Nevada University) and JUGHEAD. These provided searchable indexes of Gopher menus (Howe; 1999). Interestingly, Gopher and the World Wide Web debuted at roughly the same time in 1991. Initially, Gopher was utilised and developed at a much greater speed than its counterpart, which now enjoys domination (Howe; 1999).

Electronic-Mail

An interesting by-product of the creation of the Internet was the development of electronic mail (more commonly known as e-mail). As early as 1972, e-mail was introduced and for the following decade, it became the most popular network application (Leiner *et al*, 2000:3). This application allows researchers not just to send volumes of data via the system but also to engage in private correspondence and facilitates the exchange of ideas (Winder; 1994:2). Until the early 1990s, e-mail use remained limited to closed communities. However, the first American national commercial Internet service, Delphi, provided e-mail connection to the general public

in July 1992 and this soon changed (Howe; 1999). By 1995, it was estimated that 30 – 40 million people in over 160 countries had access to e-mail. (Elmer-DeWitt; 1996:16) "Suddenly, it seems everybody has an e-mail address from Hollywood moguls to the Holy See" (Elmer-DeWitt; 1996:17). E-mail has certainly become an integral part of millions and millions of people's lives and remains extremely popular. A recent survey undertaken by Gallup found that over 50% of respondents reported that using e-mail is their most common Internet activity. Further, 74% claim to own more than one e-mail address (Nua; 24 July 2001).

The World Wide Web

As already mentioned the World Wide Web (WWW), which today enjoys the most prolific use out of all the other Internet applications, was only developed in the early 1990s. Tim Berners-Lee, then an employee of a European Scientific Organisation (CERN) developed the WWW to facilitate easy collaboration on global scientific projects (CERN; 1998). The World Wide Web, or the W3 as it was first known (Winder; 1994:5), created an easy and consistent system that allowed scientists to access data and information in a simple and effective manner (CERN; 1998). Although slower to develop and be taken up by the public, the WWW experienced rapid growth in 1993 with the development of the first graphical browser, Mosaic (Howe; 1999). This opened up the Internet and the WWW to the general public – it was no longer limited to its previous specialised audience. The emergence of later browsers – Netscape (1994) and Microsoft's Internet Explorer (1998), consolidated the WWW's position (Howe; 1999).

Many people have hailed 1994 as the year of the World Wide Web (CERN; 1998). This was when the first major conferences were arranged to discuss the new technology and its potential. At this time and indeed even today, numerous debates exist over the usefulness, potential and future of the Internet and more specifically, the WWW.

It is important to note that the WWW and the Internet are *not* synonymous with each other. The WWW is simply one application of the Internet. Others include electronic mail (discussed above), remote data acquisition (FTP) and remote computer log-in (telnet), video-conferencing (Buchanan; 1997:1). The WWW is a combination of the computer networks and hypertext – which is “text with links to further information, on the model of references in a scientific paper or cross-references in a dictionary” (CERN; 1998). Buchanan describes the WWW as an information infrastructure where information is stored on the WWW by Web servers and employs the Internet to transmit data globally (1997:87). Information can be transmitted in various formats – anything from images to video to just plain text files. A Web page is written using the standard language, Hypertext Mark-up Language (HTML) and JavaScript. Each ‘page’ has its own unique address, known as a Universal Resource Locator (URL), and these pages can carry several links to other pages that provide further information (Buchanan; 1997:87).

Section two: How the Internet is used

It is estimated that over 409 million people regularly use the WWW for their various purposes. At the moment, those in the US outnumber those connecting elsewhere in the world, with an estimated 167.12 million Americans online. However, weekly reports inform of the ever-increasing numbers of people in all countries connecting to

the network. Nua recently reported that over 40% of Australians are online, with 23% enjoying access at home. Another report states that the number of people in China who access the Internet, tripled in 1999 (4th January 2000).

The Internet is used in ways that were never intended nor envisaged. Commercially, billions of dollars worth of annual investment is made. Businesses, small and large, have flourished and failed due to this network. The commercial sector has been involved in the Internet (then APRANET) since the early 1980's but the most notable growth took place in the 1990s. Leiner *et al* (2000:13) comment that "the Internet has now become a 'commodity' service and most of the latest attention have been on the use of this global information infrastructure for the support of other commercial services...facilitated by the WWW technology".

Despite the phenomenal growth in the 1990's with the development of the user-friendly graphical browsers (Netscape and MExplorer), many remain highly cautious of the WWW or the 'information superhighway' as it has been dubbed. Various criticisms have been levied - "It can be dangerously habit forming and if the truth be told, an enormous waste of time" (Elmer-DeWitt; 1996:21). The criticism that this technology will not enhance our lives but rather do the opposite has also been asserted. Reed Karaim asks;

"huddled in our homes, amid neglected cities, our brains numbed into a stupor while our hearts race like engines popped into neutral because of television, turning late at night to the illusory companionship of strangers in electronic worlds where those who disagree with our prejudices can be dispatched with a keystroke,

unable to relate to the leisurely reality of a sunset or trees creaking in the wind, surrounded by gadgets, yet more alone than ever – Is this really to be our future?”

(1996:126).

Although these comments were made five years ago and the WWW has developed tremendously in that time and secured a place in millions of people’s lives globally, many of the early criticisms prevail today. Indeed, it was reported last year that Rabbis in Israel have passed a ruling, banning the Internet from Jewish homes as it is “1000 times more dangerous than television” and thus vastly increases the infiltration of immoral and sinful information (Sunday Times, 9th January 2000). The Rabbis are not the only group to be concerned about the lack of control mechanisms to monitor the content on the WWW. The Communist Government in China has apparently welcomed the introduction of the Internet but attempt to censor much of the content (Nua, 4th January 2000).

It remains that people are fascinated by cyberspace, a term first coined by William Gibson in his fictional novel, Neuromancer (1984), and what it means to the world in which we live, infiltrating almost every sphere of human life (Elmer-DeWitt; 1996:16/17). As already stated, the information potential is colossal; information on virtually any topic can be obtained with relative ease, but with this technology, a new dimension of communication has emerged. More and more people turn to the Web for interaction with like-minded people. Chat-rooms, often called cyber-cafes, are now immensely popular. It seems like an infinite number of chat-rooms, on copious

topics², are available for anyone to enter. Online games in multi-user dimensions (MUDs) are profuse. There has been an immense rise in the so-called 'virtual communities'. These communities arise from the coming together of people to discuss a particular issue. This can be anything from a sporting/leisure activity³ to something more serious in one's life, for example dealing with alcoholism⁴. Often, members of these communities claim to feel a deep sense of belonging and are often very protective of their group.

When Kupfer (1996:88) claimed in 1996 that "the new technology holds the potential to change human settlement patterns, change the way people interact with each other, change our ideas of what it means to be human" he was not far wrong. We have seen that the WWW has had an incredible impact on many lives: in the commercial, information and communication domains. Whether it really will change "what it means to be human" remains to be seen.

Section three: The Internet and academic applications

As highlighted earlier, the Internet, and later the WWW, was developed to facilitate scientific (academic) work. Although the Internet and its various applications have found a huge following within the business world and with the general public, its original purpose has not been lost. Globally, the academic world is thriving owing to this invention. The Internet has already become an invaluable tool for information exchange and retrieval. Academics, within the same university and across the world,

² A search of Yahoo's chat rooms found a total of 42 categories, ranging from Business and Finance to Science. Each category held anything between 2 and 15+ chat rooms (<http://chat.yahoo.com/c/roomlist.html>)

³ For example: DBA Online: Historic War Game Community (<http://www.dbaol.com>)

⁴ For example, Sober24 an "online community and resource for persons in recovery from alcohol and other addictions as well as their family and friends (<http://www.sober24.com>)

can communicate in a fast and a convenient manner. Data becomes readily accessible. This is particularly common in the natural and physical scientific world, where it is often found that raw data (of observations or measurements) are e-mailed or posted on the Web for colleagues across the world to access and utilise (Newsweek; 11 October 1999). Global projects, with an unlimited number of participating academics, can take place without the researchers even meeting each other. Most importantly, ideas are debated with an ease and frequency that has not been seen before. There are countless discussion groups or listservers that deal with almost any subject/topic. More and more electronic journals have been created and indeed many texts, from concept papers to dissertations, are finding their way online.

However, this sharing of information is not limited to natural and physical scientists. Academics from all subjects can be found to be engaging in similar exchanges – and the Social Sciences are no exception. The information resources for the Social Sciences are immense, on topics ranging from sociological theories⁵ to the pages assisting you with the finer details of statistical analysis⁶.

Adjacent to this, the Internet has opened up many areas of study for Social Scientists. Documentary material is a primary source for all researchers. Personal documents, official and unofficial reports, records, case-history material, research findings, theses, papers, etc. are often nowadays posted on Internet. A vast number of Social Science journals are now available online -from those placed online in addition to traditional

⁵ For example: Sociosite holds a list of websites dealing with various theorists and theories (<http://www.pscw.uva.nl/sociosite/topics/theory.htm>)

⁶ For example: SelectingStatistic: a program developed by Cornell University to aid in the selection of the right statistical test, hosted on Bill Trochim's website (<http://trochim.human.cornell.edu/selstat/ssstart.html>)

print, to those published solely on the Internet⁷. Numerous databases are available on the WWW, facilitating secondary analysis of data on widely varying subject (and geographical) areas⁸.

In terms of primary data collection it is argued, (Lee; 2000:118) that the network opens up previously inaccessible areas of study. For example, studies involving people living in rural and remote areas, members of various groups (often those that are unusual or deviant) and cross-national studies become more feasible with the use of the Internet. It is suggested that psychologists can transfer some of their experiments onto the Internet, with the benefit of reaching more potential participants than otherwise would be found (Lee; 2000:117).

Sociologists and anthropologists are also using the net for primary data collection purposes – quantitative research (e.g. surveys) on the World Wide Web is now not unusual and the use of it for qualitative (ethnographic) research is on the rise. This is not surprising considering the expeditious growth of the WWW and its integration into the lives of many ordinary people. The whole notion of cyberspace and the many thousands of communities that have arisen as a result provides social scientists many new areas of study.

There exists a substantial quantity of research that has been undertaken using the Internet as the main means of data collection. Many groups, ranging from

⁷ For example: Social Science Computer Review, traditional print journal (<http://hcl.chass.ncsu.edu/sscore/sscore.htm>) Sociological Research Online, which is only available on the Internet (<http://www.socresonline.org.uk>)

⁸ For example: UK Data Archive, maintained by the University of Essex (<http://daww.essex.ac.uk>) or The Question Bank, maintained by the Centre for Applied Social Surveys at Surrey (<http://www.scpr.ac.uk/cass>)

communication analysts, behavioural scientists to those interested in the social impact of the computer technology, are now undertaking studies. A look at a site such as <http://www.netzwissenschaft.de/> (a German site, containing sufficient English to assist non-German speakers) presents a rather large list of on-going studies in a number of fields. Thus it is evident the Internet is being used as a research tool.

We have established that the Internet is being utilised by Social Scientists and the different ways in which they are doing this. However, this use is not problem-free. In this section, I will mention the problems that are faced by using the technology in the various ways and, more importantly, turn to the research problem addressed in this study.

The first problem concerns authenticity of the information, be it in the form of articles, reports and other documents that are found on the Internet. The authenticity of documents obtained this way is often in question and means of verifying their authenticity may prove to be more complex in the case of Internet documents than hitherto experienced. One reason for this is owing to the fact that the sources are often so geographically distant that corroboration is difficult. Another problem is the plethora of sources (documents, records, and reports) that become available to the researcher at the click of a button. Other problems that arise with this information super-highway include issues of copyright and the ease of plagiarism. These issues, and many more pertaining to the information available via the Internet, remain unresolved. Steps can be taken in an attempt to verify the authenticity of a

document⁹ but is by no means foolproof. However, a detailed examination of these problems falls outside the scope of this study, to which we now turn.

It is evident that various social scientists are employing the Internet for primary data collection purposes. Thus, it is important to ask if the established desiderata of acceptable research in the social sciences are being met. What methodological questions are raised by this new means of data collection? Do the tried and tested methods suffice or are adaptations to the methods required? For instance, when undertaking empirical research it is important that the correct sampling procedures are followed or the established interviewing techniques used. These standard methods, if closely followed, assist in minimizing error for the attainment of valid and reliable results in conventional research. But is this the case for Internet research? To stress the questions I want to address: Can the established methods be transferred to the Internet? Is the new research undertaken producing valid and reliable results? Does research, using the Internet for data capturing purposes, require a major change in research design?

Concluding comments

In this chapter, we have seen that the Internet has quickly developed into a global computer network, from which many innovative applications have evolved. The most popular, and perhaps the simplest, is electronic mail (e-mail). It allows people to send messages and documents to one another with utmost ease and speed. Another highly publicised feature of this new technology is the World Wide Web, which provides

⁹ See for example Sosig's Internet Detective (www.sosig.ac.uk/desire/Internet-detective.htm), 'an interactive tool', on the evaluation of Internet Sources.

rapid access to information though out the world and provides a new form of communication.

This new technology brings another dimension to the study of societies. It is hoped that the systematic enquiry into social life will be greatly enhanced by using the technology. Information; articles, reports, thesis, are often disseminated via this medium. Alongside this, communication on the Internet opens up a new era of academic dialogue.

Empirical data collection using this medium is becoming more common. Quantitative data can now be obtained globally, by means of an online survey providing opportunities for mass data collection on social phenomena unseen before. Researchers can also collect qualitative data, again in remote areas but also particularly concerning sensitive case studies, previously difficult to obtain. Finally, many studies, of both qualitative and quantitative nature, have been undertaken on the Internet about the Internet, for example investigations into the many online communities that now exist. These issues bring us to the main focus of the study - to address the methodological questions (highlighted in the previous section) raised by such studies, beginning with the following chapter, which is an examination of survey research using the Internet as a medium for data collection.

Chapter Two

A CLOSER LOOK AT SURVEY RESEARCH ON THE INTERNET

Introduction

Questions need to be asked about how long-established quantitative methods, specifically the methods of survey research, hold up when employing the Internet as the medium for data collection. By survey research, we mean research that

"exposes a group of people representative of a target group, to which the researcher expects to generalize, to common situations or stimuli and records their reactions. The researcher is usually interested in their common responses to the questions, the variability in responses, and the interrelationships of certain responses, especially those involving demographic information or measures of social or psychological variables with positions on issues"

(Babbie; 1998:361/2).

Over the years, when using this method, strategies have been developed at the various stages to minimise bias and error. In this chapter, we are interested in whether these strategies carry over when the Internet is employed as the primary source of data collection. If they do not, can we adapt the existing strategies to the Internet or do we need to develop new ones?

There are two recognized modes of gathering data using the Internet: e-mail and the World Wide Web. E-mail survey methodology is purported to be essentially similar in nature to that of postal (mail) surveys. Various researchers (Bauman *et al*, 1998;

Schaefer & Dillman, 1998; Tse *et al*, 1995) have shown that many of the characteristics are comparable. This parallel applies, in particular, to the obstacles encountered when conducting a postal and e-mail based survey. The shared difficulties include low response rates, self-selection of respondents, little control over response quality etc. Besides these difficulties, it seems that e-mail surveys face a number of additional obstacles. Web-based surveys are a much more recent innovation and thus present many more methodological challenges. The aim of this chapter is to examine existing research, in which data has been collected by means of Internet surveys, and to consider the methodological advantages and limitations of two main aspects of Internet survey research: sampling issues, which includes non-response, and questionnaire design. Issues concerning the validity, reliability and generalisability of the research will be addressed. Furthermore, consideration will be given to suggestions as to how the identified constraints may be overcome.

Section one (A): Sampling

The proper use of sampling techniques has, over the years, consistently proven to be a necessity for the production of sound survey research. An oft-cited example of poor sampling techniques is the notorious case of the 1936 Literary Digest poll, which despite receiving responses from over two *million* voters wrongly forecast that US presidential candidate Alfred M. Landon would win convincingly over Franklin D. Roosevelt. Sample size was not the issue but rather lack of adequate representation of different members of the sample population (Babbie; 1998:193). This inadequacy serves as an illustration of the importance of following correct sampling procedures.

Shipman defines sampling as "a systematic way of choosing a group small enough to study and large enough to be representative" (1988:52). The key term here is "representative". The sample needs to adequately reflect the population in question. However, it is worth noting that not all researchers believe that representativeness is an achievable goal. Henry suggests that, as there is no set standard informing us of what constitutes a representative sample, this is subjective to the researcher. Thus, it is more correct to refer to a sample as a "model or subset of the population that is used to gain information about the entire population" (Henry; 1998:102).

Regardless of the semantics involved, it is clear that the underlying principle of sampling dictates that in order to produce valid, reliable and generalisable results, the researcher must ensure not only that sufficient numbers are used, but also that those people chosen reflect the population at large.

Coomber suggests "the key issue that any survey research conducted via the Internet will have to contend with, as with non Internet based surveys, is that of sampling bias" (1997:1). In order to fully appreciate the importance of following correct sampling procedures and the difficulties involved in doing so, it is first useful to outline the various established sampling techniques, which are used in traditional methods of survey research, before addressing issues of sampling when conducting an online survey.

Sampling methods can be divided into two main categories: probability and non-probability methods. Probability sampling is often considered as the most reliable means of compiling a sample of the population to be studied. According to Babbie,

the basic principle behind probability sampling denotes that a "*sample will be representative of the population from which it is selected if all members of the population have an equal chance of being selected in the sample*" (author's italics) (Babbie; 1990:71). This method eliminates more potential bias than non-probability sampling as well as enables the researcher to hypothesize on the accuracy of the sample - via sampling error statistic and confidence level / confidence error statistics¹⁰.

Probability Sampling

There are four types of probability sampling, namely simple random sampling (SRS); systematic; stratified and multi-cluster sampling. Before undertaking a brief discussion outlining the main features of the above sampling types, it is first necessary to consider the issues involved in compiling the sample from a population. It is worth noting that this issue has proved to be particularly problematic in Internet surveys. However, this problem will be addressed later in the chapter.

A **sampling frame** refers to the list of all members of the population to be researched from which the sample is selected (Arber; 1993:70). Ideally, this list should be complete, up-to-date and accurate. For example, if you were studying a university's entire undergraduate population, a list of all registered students would be available from student records, should you be able to negotiate access. This list would constitute your sampling frame. However, it is possible that a complete sampling frame might not be available for the population in question. In this event, it is necessary to compile the list from various sources - which may prove to be laborious.

¹⁰ See Babbie, 1990 or de Vaus, 1996 for a comprehensive discussion on these statistics.

Even if a sampling frame appears to be complete at a first glance, there may be people who have been omitted. Arber gives the example of a list of employee's of a multi-national company. It is likely that a large company will have a frequent staff turnover. Thus the list may not be entirely accurate; most recent new recruits may not appear on the list or internal transfers not noted (Arber; 1993:80). It is important to be aware of these possibilities and take steps to ensure these potential biases are covered. As Neuman comments, a good sampling frame is imperative as "a mismatch between the sampling frame and the conceptually defined population can be a major source of error" (Neuman; 1997:209).

As already stated above, probability sampling consist of four types. The first type, **simple random sampling** (SRS), provides the basis for the other three types, which have increasing degrees of complexity. The basic requirement in undertaking SRS is obtaining/compiling a complete sampling frame. Next, a decision must be taken on the sample size. This will be determined by factors such as population size, degree of confidence level desired and the number of variables to be studied in that population. In addition, it is likely that costs may be a consideration (see de Vaus; 1996:62/5). Finally, based on the sample size decided earlier, you must randomly select cases from the sample. This can be done with the aid of tables / lists of random numbers (e.g. see Neuman; 1997 Appendix B). However, there are various computer programs, such as SPSS, which have the function to do this.

Systematic sampling follows much the same steps. However, instead of randomly choosing numbers from the sampling frame, the researcher must calculate a sampling interval - i.e. every x person will be chosen for the sample. The biggest drawback of

this method occurs when the sampling frame is listed in a particular manner. Neuman (1997:211) gives the example of a sampling frame organised by married couples; male listed first, female second. In this case, if your sampling interval were an even number, only people of the same sex would be chosen, thus creating an unacceptable bias in the sample and, in this instance, SRS would be a more appropriate option.

Stratified sampling is best used when you wish to make sure that particular characteristics of your population are equally represented. This includes features such as sex, age and race. The researcher must first divide the population into the categories (strata) and then carry out SRS / systematic sampling on each of the strata. In general, this type of sampling provides a more representative sample of the population at hand.

Cluster sampling is the most complex of the probability sampling methods, most often used when a good sampling frame for a diverse group is absent and the cost to reach the sampling element is very high (Neuman; 1997:214). Thus, researchers use a sampling design that involves multiple stages and clusters. The researcher first randomly samples clusters (e.g. a sample of all colleges in a country) and then randomly sample elements from within the selected clusters (i.e. each college selected in the first sample) (Neuman; 1997:214/5)¹¹.

Non-Probability Sampling

Non-Probability sampling refers to a sampling method that is *non-random* in nature,

¹¹ For a much more comprehensive account of cluster sampling, see Babbie (1998); de Vaus (1996) or Neuman (1996)

that is, no mathematical procedure can be undertaken to determine a degree of confidence in the results found by the researcher (Shipman; 1988:57). It is precisely this lack of mathematical security that causes many commentators to be highly critical of the method. Neuman states; "a researcher uses (non-probability methods) out of ignorance, because of a lack of time, or in special situations" (1997:204), thus, indicating that these methods should be used sparingly. Nonetheless, non-probability sampling methods are at times unavoidable when employing probability methods is not practical or perhaps even possible. For example, it is possible that the population to be studied may be greatly scattered geographically (thus making cluster sampling ineffective). Another instance when probability sampling is not possible occurs where a sampling frame of the population is not available (de Vaus; 1996: 67). This is probable when studying issues of a sensitive nature or deviant groups.

Convenience sampling simply entails that the researcher select cases that are available. Obviously this implies a serious lack of representativeness, although this method can return a large number of respondents (for example television phone-in polls). De Vaus suggests this method is useful in exploratory research or pilot testing (1996:69). However, Neuman states, "the bias and systematic errors that easily occur make them worse than no sample at all" and is against their use (1996:205).

Quota sampling involves compiling lists of particular characteristics (e.g. particular age group, gender) from which the researchers must obtain a predetermined number of cases from each characteristic. "The quotas are organised so that in terms of the quota characteristics, the final sample will be representative" (de Vaus; 1996:68). This type of sampling remains non-random, as interviewers can select (or fail to

select) any person who meets the criteria and can simply keep going until the quota has been filled. Furthermore, accurate population proportions may not be available, thus, the researcher only estimates the representiveness of the sample (de Vaus; 1996:69). The case of the 1936 Literary Digest poll, which wrongly predicted the successful presidential candidate, is an illustration of how quota sampling can produce grossly misleading results (Babbie; 1998:196). However, a number of mechanisms exist (ranging from simple to more complex and hence more costly) that can be utilised, which offers some control over the procedure (Arber; 1993:75/6). Arber also asserts that a researcher is able to evaluate the representativeness of a quota sample by "comparing the sample with known population characteristics, using characteristics other than those specified in the quota controls" (1993:77).

Purposive sampling involves the use of a researcher's judgement - a case (area / school) may be selected, as it is perceived (judged) by the researcher as being 'typical' of its kind (de Vaus, 1996:68). Neuman (1996:206) cites three cases where purposive sampling is the most suitable method: the selection of unique cases which are particularly informative; the selection of cases in a difficult to reach, specialized population; and its usefulness in the identification of cases for in-depth investigation. Shipman warns of the hazards of employing purposive methods, especially in the case of volunteers, "(it) is right to suspect that there is something unusual about those who step forward when a researcher asks for help" (1988:58).

The last type of non-probability sampling is referred to as **snowball sampling**, also known as referral or reputational sampling. This method is useful when a researcher is interested in studying a network of people. It starts with a small number of people,

who are interviewed and then asked to refer other people of similar characteristic - thus the researcher ends with a large network of interconnected participants (Neuman; 1996:207). Snowball sampling is also useful in the absence of a sampling frame. The method works in the same way as described above but instead of the interest in the network itself, it is simply a means to identify people with the desired characteristics. In particular, it is useful in the study of deviant groups - for example, those involved in illegal activities (Arber; 1993:74). Although the pitfalls of this method are clear, the researcher can only tap to that particular connected network, thus, potentially creating a grave source of bias (Arber; 1993:74).

The above section has outlined, albeit briefly, the concepts behind both probability and non-probability sampling. It illustrates the concerns that researchers are required to take into account in the endeavour to reach the research goals of generalisable findings. The issue, which must now be addressed, examines how surveys, employing the Internet as the main means of data collection, measure in the compilation of a sample.

As stated in the introduction, there are two accepted means of collecting data on the Internet: by means of e-mail or the WWW. (It should be noted that these methods can be, and often are, used simultaneously for optimum results.) The application of traditional sampling methods and the problems that arise with this will be explored in the following section. For the sake of clarity, throughout this chapter, I will address e-mail surveys and WWW surveys independently.

Sampling issues in e-mail surveys

The first major obstacle to overcome when undertaking an e-mail survey is that of compiling a sampling frame. This can prove to be exceptionally problematic. The new and ever-changing nature of the Internet means that a sampling frame of the population simply may not exist and difficulties may arise in putting one together.

The ownership of an e-mail address is not pervasive in the same way as a telephone number. The widespread possession of home telephones has meant that telephone directories are perceived to be an adequate tool, in certain situations, for use in compiling a sample frame, although it is important to recognise that they are not regarded as complete. Not everyone has a telephone and of those that do, a number may choose to be unlisted (often those in the higher income bracket). Contrastingly, those people who have an e-mail address and who regularly make use of it, are most likely to be well educated and subsequently belong to a higher income group. As discussed in the first chapter, research indicates that the majority of the people, who frequently use the Internet, belong to the middle and upper income groups. Thus, it can be deduced that an e-mail listing is likely to eliminate those in the lower income-bracket and thereby introduce a serious class or socio-economic bias into the sample.

As of yet, there is no comparative (or "comparable") directory of e-mail addresses. Certainly, e-mail listings do exist. For example, most of the major search engines offer a 'people' search or directory. However, there is a long way to go before they will be viewed as enjoying the same comprehensiveness as telephone directories. Furthermore, telephone directories are recorded via region. Therefore, a researcher is simply able to determine from which area he or she would like to draw a sample

population. This is not the case with e-mail directories. Regional information may be available, but not always, and it is likely that it consists of people from all around the world. For example, if a researcher were interested in surveying all the residents of a particular city (i.e. Cape Town) it is not yet feasible to do this by means of e-mail. A great number will simply not own an e-mail address and of those that do, it would be very difficult, if not impossible, to attain a list of all their addresses from these directories.

Another consideration, which has to be taken into account with the use of telephone directories, is the number of households that move each year. This fluctuation implies that a telephone directory quickly becomes inaccurate (Lavrakas; 1993:48). A similar problem exists with e-mail lists. Due to the transient nature of the Internet, e-mail addresses are liable to change rapidly (Sheehan & Hoy; 1999:2).

Nevertheless, a growing number of situations exist where e-mail surveys are not only feasible but perhaps the best option. Within the work environment, e-mail is pervasive. Most sizeable institutions use the Internet, e-mail in particular, as an essential part of their daily operation. Consequently, an e-mail survey of such an institution is not only possible but also advantageous in a number of ways. A directory of employees and e-mail addresses (and thus sampling frame) are likely to be more readily available from say, the Human Resources or IT department. However, even if this is not the case, it is likely that a complete list can be put together without exceptional difficulty.

An illustration of this is Schaefer and Dillman's (1998:379) survey of an American higher education institution. Existing data on the permanent staff of the university indicated that 89 percent had access to an e-mail address. Hence, Schaefer and Dillman (1999:382) contend that the "coverage error from an e-mail survey, while higher than desirable, might be acceptable". In this study, the authors' main concern was a reduction in non-response rate and so the study and its methods are discussed in greater detail in the following section.

Another survey using employees of a university was undertaken by Tse *et al* (1995). In this study they employed the "Telephone Directory of the Chinese University of Hong Kong", which contained both e-mail and postal mailing addresses of academic and teaching staff at the University. 400 names were picked (although the authors do not comment on how this number was reached and the procedure in choosing these 400 names). The above two examples illustrate the ease in which an adequate sampling frame can be obtained within the confines of an institution and, in this instance, established sampling procedure can be followed.

More difficulties in identifying a sample population arise when not studying a closed group (such as an institution). Sheehan and Hoy (1999) set out to investigate the potential of e-mail in surveying Internet populations both nationally (USA) and internationally. The authors drew a sample population from "known Internet users in the United States who access the Internet for their personal use" (1999:6). As no single e-mail directory exists, the authors were compelled to be innovative and in this process developed a rather complex stratified sampling procedure.

The e-mail sample was compiled by using the Four11 Directory Service, which allowed them to use the search field to generate e-mail addresses by using the 'domain name' field (i.e. abc.edu). On average, this would bring up around 200 addresses per domain. 55 randomly selected domain names of Internet Service Providers (ISP) were compiled by using Network USA's Internet Service Provider Catalogue, the List and the Internet Access Provider Meta-List (these are all lists of commercial personal ISP's in the US developed by ISP submissions for inclusion) (1999:7). A search, using Four11, of the selected domain names was then undertaken. This showed all addresses up until 200, thereafter only a sample of 200 of the total addresses were displayed. However, the total number of matches was provided and from this the authors developed a system for selection. If there were less than 1000 matches, every fourth name of the 200 name sample was selected. If there were more than 1000 matches, then all of the 200 in the sample would be selected. The names and addresses were manually selected and added to a mailing list.

As the authors point out, the ability to create a search using the domain field was particularly useful. Unfortunately, since the data collection for the survey was undertaken, Yahoo! bought over Four11 and this facility no longer exists. Regrettably, none of the other people finders offer this facility either (1999:8). However, this study has shown that it is possible to compile a sample for an e-mail survey.

Within all research, a sample population and sample frame is entirely dependent on the topic to be studied. The above case was effective in producing a reasonably representative sample as it focused on the perceptions of *current Internet* users. However, the method would not have been considered in any way representative nor

the result in any way valid or reliable had the study been concerned with a subject matter such as the American population's view of the Internet (it would have excluded non-users and thus including an unacceptable bias).

As highlighted during the discussion on sampling methods, it is not always possible to compile a sample via probability methods. Coomber (1997:6.2) decided to conduct online research on illicit drug dealers and as a result had to rely entirely on non-probability methods. As a way of targeting members of his population and inviting them to participate in his study, he conducted a search to identify drug-related newsgroups, which returned 23 main groups. Requests for dealers to participate in the study were posted to the majority of these newsgroups. This method was successful in reaching a number of people from the study population. However, owing to the nature of Coomber's study, potential respondents were apprehensive that anonymity could be assured (an issue which will be returned to at a later stage), and securing respondents was another task altogether.

The above discussion has highlighted the fact that although the technology involved with an e-mail survey is far removed from either mail or telephone based surveys, and that additional sampling problems are encountered, the concepts involved are not that different and further mechanisms exist that allow the compilation of an acceptable, representative sample. However, the World Wide Web is essentially different in nature and it is foreseeable that many more difficulties could arise.

Web-Based Surveys

At first glance, the question could be asked whether a researcher can have any control over the respondents of a Web-based survey at all or at best suggests that it

possesses the potential for nothing more than convenience sampling. If this is the case, the likelihood of sampling bias is even greater with web-based surveys than with an e-mail survey, owing to the non-random nature of non-probability methods. This suggests that web based surveys, although possible will not produce representative (and hence non-generalisable) results. However, a number of researchers have undertaken such studies and offer potential solutions to this problem.

All surveys, be it interview based, telephonic or mail, invite prospective respondents to participate in the research. In these methods, this is done after targeting the sample population and taking steps to make contact. A web-based survey also invites people to respond. However, the researcher has less control over those that actually choose to participate, as individual solicitation is not normally undertaken. Thus, a number of steps need to be employed in order to maximise contact with the target population and entice as many as possible to participate.

Bradley (1999:3) suggests there are a number of ways to solicit survey respondents. Announcements/invitations (with or without a hypertext link to the survey website) can be strategically placed in relevant websites, discussion groups, printed magazines/newspapers etc. Alternatively, e-mail addresses can be harvested from directories (either ISP's or Internet e-mail services, e.g. Yahoo! Mail) customer/staff records or subscribers to discussion groups/websites. These e-mail addresses would be then be used to send individuals a request to participate in the survey.

A number of these methods are employed by Kehoe and Pitkow (1996) in their GUV (Graphics, Visualisation and Usability) WWW User Surveys, conducted annually. For example (ibid; 1996:3):

- Links to the survey from high-exposure, general-interest Web sites, such as NCSA's "What's New", Yahoo, Lycos, CNN etc.
- Announcements on WWW and Internet-related Usenet newsgroups
- Coverage in national and local newspapers and trade magazines
- An announcement on the www-surveying mailing list that we maintain for users who would like to be notified about GUV survey activities.

The authors (ibid, 1996:3) felt that promoting the surveys through various media would increase the potential of reaching a larger number and more diverse group of respondents. Further, in order to find out the most successful means of 'advertising' of their study, they included an item in their questionnaire inquiring as to how the respondent became aware of the survey. This information proved useful when planning their subsequent surveys.

Three different types of web page questionnaires have been identified: open, closed and hidden (Bradley; 1999:4). As the name suggests, open-web questionnaires are available for anyone to complete. This method holds little control in terms of those completing the survey. However, this lack of control is not necessarily negative - an example of this is found in Survey2000 discussed below. A closed-web questionnaire refers to a web survey in which potential respondents are invited to go to a particular site to complete the questionnaire. It is likely that this site would be password protected. Finally, a hidden web questionnaire refers to those that become visible only after they have been triggered - an example of this is a 'pop-up' survey. At a chosen web site a random sample of visitors are offered the opportunity to complete

the survey. The researcher is able to determine the frequency of the invitation - be it per number of visitor (i.e. every fourth person) or over a fixed time frame. There is evidence that this method is a more effective means of obtaining respondents than a fixed banner invitation, although the measurement of web-site traffic is not always reliable and the researcher should be aware of this when considering the possible employment of this method (Bradley; 1999:2/3).

Witte *et al*, in the Survey2000 Project, undertook an interesting example of an open-web survey. This project is currently the largest and most comprehensive social survey on the Internet. It looked at mobility, community and cultural identity (Witte *et al*, 2000:179). The survey comprised of three different forms: the Canadian and U.S. Adult Survey, the Youth Survey and the International Respondent Survey (respondents were asked to chose the appropriate form to complete) (ibid. 2000:180). As the project was undertaken in collaboration with National Geographic (NG), a link to the questionnaire was available on the NG web site for a two-month period. This link allowed any person who visited that site the opportunity to take part in the survey. Further, the survey and web site address was publicised via advertisements (even before the survey went online) in National Geographic's adult and children periodicals, television and school based education activities (2000:189).

The authors comment on the non-random nature of a web survey. They made use of two mechanisms in order to overcome this difficulty. Firstly, and perhaps most credibly, items from existing (traditional) surveys (e.g. GSS) were used to "provide external benchmarks to assess the nature of the survey bias and construct the necessary weights" (2000:185). Secondly, the extensive use of National Geographic's

public relations and community outreach resources were considered to play a significant role in increasing awareness of the survey as well as providing credibility in the public eye (2000:185).

The use of other surveys to provide information in assessing the survey bias proved to be very useful and interesting. The authors found that the Internet gender gap has all but disappeared in the United States. However, 92% of respondents were white, which is indicative of the under-representation of minorities on the Internet. Finally, a large disparity in educational level was found between the Internet survey respondents and the existing survey material. Nevertheless, with the aid of the existing survey material these problems can potentially be overcome, e.g. by weighting (a standard measure to deal with deviations from equal probability sampling)¹², which would then allow generalisations to be made.

Despite these measures, the authors commented, "the potential for sample bias in data collected from the Internet represented the most serious methodological problem facing Survey2000" (2000:118). In this survey, the probabilities of sample selection and even the size and boundaries of population membership are unknown. This refers to who is online and who visits the National Geographic site. Hence, it raises the issue of representativeness; how many participants are required in order to obtain a representative sample. The total number of completed Survey2000 questionnaires amounted to 54,937. Nevertheless, the notorious Literary Digest poll, referred to earlier in this chapter, is once again a good illustration that a large sample

¹² Neuman (1997) or de Vaus (1996) provide a comprehensive, detailed discussion on the concept of weighting and its application.

size is still not a guarantee of representativeness. It is the authors' view that the use of "a large voluntaristic sampling approach, along with external benchmarks and an aggressive outreach effort to diversity the sample" was reasonable compensation for the deviation from random sampling (2000:192).

In conclusion, the authors postulate that the un-weighted Survey2000 data is likely to produce considerable insights into mainstream web users. They observe that it appears to be possible to construct plausible weights, allowing for large-scale generalisations (of the US population), although the "selective nature of sample needs to be taken into account" (2000:192).

Section one (B): Non-response

As the above section has shown, the identification and compilation of a sample is a difficult task, in both traditional and Internet surveys. However, the successful completion of this stage does not ensure that the data obtained is unbiased. A big threat to the validity and reliability of data collected comes in the form of non-response. All types of surveys are subject to this problem, albeit to different degrees. Non-response presents two potential problems, namely a dramatic reduction of sample size and the possibility of bias (de Vaus; 1996:73). There are various techniques that can be employed in order to deal with the reduction of the sample size (this will be dealt with in detail later in this section). On the other hand, the problem of bias is more difficult. Often little is known about those people who have declined to partake in the study. It is possible that they possess certain characteristics that differ from those who do complete the survey. That these people are not partaking in the study might mean that the data is not representative of the

population as determined in the original sample. Further, as illustrated previously, increasing the sample size will not necessarily increase representativeness. De Vaus observes that while there are statistical techniques that can be employed to minimise the effect of bias during analysis of the data, it can be difficult to uncover the bias and the degree to which it occurs (1996:73).

It is clear that a higher response rate lowers the possibility of response bias. Various commentators differ on an acceptable response rate. Mangione considers a response level that is between 60% and 70%, acceptable (1995:60). However, according to Babbie, (and generally accepted) 50% is an adequate response rate in mail surveys (1998:262).

Non-response and/or refusals to complete a questionnaire occurs for various reasons: the respondent could not be contacted (i.e. wrong address) or, if contacted, did not have time, was ill, does not speak the language of the questionnaire. Another possibility is that the respondent is uncomfortable with the survey topic or a number of the questions posed (Neuman; 1997:247). According to Mangione, who analysed information from ten surveys to build a profile of these characteristics, non-responders are likely to be "less educated, or elderly, or unmarried, or male, or to have some characteristic that makes them seem less relevant to the study" (1995:61). It can be helpful to try and obtain information on non-respondent characteristics for later comparisons with respondent characteristics. De Vaus suggests that a certain amount of information can be collected merely by observation; for example, in a face-to-face situation, the age group, sex and ethnic background of

the non-respondent can be determined. It may even be possible to acquire details on housing, car, neighbourhood etc (1996:74).

Another means of obtaining information on non-respondents can be gathered from the sampling frame. This is especially pertinent when surveying the employees of an organisation or institution. Human resources may be able to provide general information for the entire sample population, thus the researcher is able to determine whether there are any set characteristics of those declining to complete the survey data (de Vaus; 1996:74). Voter registration lists also provide a selection of demographic details. By keeping track of who has and has not responded, this allows certain inferences to be made about non-respondents (Mangione; 1995:61). In postal surveys, a means of obtaining information on non-responders is to send a letter requesting some demographic information. Mangione employed this method when undertaking a survey on alcohol usage. A one-page letter, asking for a number of demographic details, was sent to non-respondents to complete. A \$2 monetary incentive was also included. This request received a 50% return from these non-responders, providing valuable comparative information (Mangione; 1995:61).

Techniques for reducing non-response error in traditional surveys.

The reduction of the sample size to unacceptable levels, as a consequence of non-response, generates biased data. A number of studies have focused on the problem of non-response in surveys, which has led to the development of several techniques that are valuable in reducing non-response rates.

In the case of face to face interviews, it has been suggested that an advance letter or alternatively a phone call (especially when studying those from inner-cities and lower-income groups) explaining the nature of the survey, together with a request to participate, has proved to be successful in reducing refusal rates. Prior to commencing the interview, the respondent should be fully informed on the subject matter, the importance and legitimacy of the study (Arber; 1993:84). Further, it is recommended that the interviewer carry photographic identification to increase respondent confidence in the validity of the study (Neuman; 1997:247). If the respondent is not at home, a certain number of repeated visits should be made at varying times on different days.

Measures for telephonic questionnaires are similar to those in face-to-face interviews. A covering letter preceding the telephone call has proved to be more than effective in increasing response rates. A similar introductory speech as recommended for face-to-face interviews, should be made at the beginning of the call. Should the respondent not be available, it is recommended that several call-backs be made at alternative times on different days. Lavrakas states that for academic and government sponsored research, anything between 10 - 20 call-backs are planned (1993:88). However, Neuman deems five to be sufficient (1997:247).

Of all the traditional methods, mail surveys encounter the most difficulties with non-response. Indeed, many papers and chapters are devoted to this problem and a number of preventative measures have been suggested (see Dillman, 1991; Fowler, 1984; Rea & Parker, 1992). As mentioned earlier, e-mail surveys are very similar to postal surveys; therefore, it is pertinent to examine the approaches developed to

tackle non-response in detail. It is essential that the questionnaire should be accompanied by a covering letter (preferably on official letter-headed paper) explaining the study, sampling procedure and assurances of anonymity / confidentiality guaranteeing that, should they participate, they will not be able to be identified in any way (Mangione; 1995:63/4).

The most obvious way of reducing non-response rates is to send out "follow-up mailings" (Rea & Parker; 1992:85). Mangione recommends the first thank you / reminder postcard should be posted 14 days after the initial mailing. The second follow-up mailing should be posted 14 days after this (Mangione; 1995: 66/7). Should the project budget allow, a third follow-up letter yields good results. It is also recognised that the provision of pre-paid postage for the return of the questionnaire is valuable in enhancing response rates. Incentives (monetary or otherwise) are another established means of increasing response rates, although not all research budgets are able to cater for this. In this event, a number of researchers have used creative ways to overcome their budget constraints. Babbie recounts some of these methods: a group of researchers included two cents with their questionnaire stating "we want to get your two-cents' worth on some issues and we're willing to pay"! Others have included a quarter (around R2) and said "make a child happy" (Babbie; 1998:262).

Research has also established that the design and length of the questionnaire dramatically affect response rates. The shorter the questionnaire (four pages or less) and a clear, easy to understand physical appearance, generally increase the number

of responses (Neuman; 1997:247/8). However, issues on the design, format and appearance of the questionnaire will be addressed in detail in a later section.

Non-response in e-mail surveys

The difficulty of gaining adequate response rates is as much a problem with e-mail as with postal mail surveys. It is only in recent years that researchers have started to focus on issues of non-response in e-mail methodology. Interestingly, the research has yielded conflicting results. Tse *et al* reports on the first comparative study, undertaken by Schuldt and Totten (1994), between these two methods. This study found that the e-mail response rate was 'pitifully low' (1995:441). As mentioned during the discussion on sampling, Tse *et al* undertook a similar investigation to Schuldt and Totten. In this survey of administrative and academic staff at the Chinese University of Hong Kong, a return of 27% for the postal questionnaire and 6% for the e-mail questionnaire was obtained. The difference was statistically significant. Interesting, they also tested for differences in response speed and quality, neither of which returned statistically significant results (Tse *et al*; 1995:444). As a caution to these findings, it is important to take into account the age of this study. During the past six years, e-mail has become far more entrenched in people's everyday lives and become more of a universal phenomenon. Despite this fact, the Internet remains very fluid in nature, and as a result, ever changing. As mentioned earlier, this means that e-mail addresses change frequently. Hence, it is likely that a number of the e-mail address in a sampling list will prove to be incorrect, presenting real difficulties for the researcher.

The following table (adapted from Sheehan & Hoy; 1999:11) lists the response rates of 11 surveys, which used multiple contact methods (mainly postal mail and e-mail) to approach respondents. The table shows the variety of studies undertaken on the topic and clearly illustrates the conflicting levels of success that researchers have enjoyed with e-mail response rates, including Sheehan & Hoy's own study (discussed in the section on sampling methods), which produced a final response rate of only 24%.

Author	Response Sample	Survey Topic	Method	Sample Size	Usable Sample	Response Rate
Kiesler & Sproull (1996)	Employees of a Fortune 500	Corporate Communication	Mail E-mail	115	77	67%
				115	86	75%
				230	163	
Parker (1992)	Employees of AT&T	Internal Communication	Mail E-mail	70	27	38%
				70	48	68%
				140	75	
Schuldt & Totten (1994)	Marketing & MIS Professors (US)	Shareware Copying	Mail E-mail	200	113	56.5%
				218	42	19.3%
				418	155	
Mehta & Sivadas (1995)	Usenet Users	Internet Communication	Mail E-mail	309	173	56.5%*
				182	99	54.3*
				491	272	
Tse, <i>et al</i> (1995)	University Population (Hong Kong)	Business Ethics	Mail E-mail	200	54	27%
				200	12	6%
				400	66	
Bachman, Elfrink & Vazzana (1996)	Business School Deans	TQM	Mail E-mail	224	117	52.5%
				224	147	65.6%
				448	264	
Sheehan & Hoy (1997)	University Population (Southeast US)	Privacy and New Technology	E-mail	580	274	47.2%
Smith (1997)	Web Presence	Business Activities	E-mail survey E-mail solicit.	150	11	8%
				150	42	11.3%
				300	53	
Schillewaert Langerak & Duhamel (1998)	Web users in Belgium	Attitudes towards the Web	E-mail Ad in mag. USENET posting Hyper links	430	125	31%
				62.5M	110	0%
				4000	67	2%
				7500	51 353	0.68%
Weible & Wallace (1998)	MIS Professors (US)	Internet Use	Mail Fax E-mail Web Form	200	70	35.7%
				200	50	30.9%
				200	48	29.8%
				200	52	32.7%
Schaefer & Dillman	University Faculty	Unknown	Mail E-mail	226	130	57.5%*
				226	131	58.0%*

*Differences are not significant

(Adapted from Source: Sheehan & Hoy; 1999:11)

Schaefer and Dillman's (1998:379) survey used a multiple contact, mixed mode method of both e-mail and postal mail in order to develop and test procedures for conducting e-mail surveys. As with many studies, the authors used university staff members (permanent faculty of Washington State University) as their study population. The 904 staff members were randomly divided into four groups, each of which received four contacts as shown in the table below:

	Group 1 All Paper	Group 2 All E-mail	Group 3 Paper Prenotice	Group 4 Paper Reminder
Prenotice	Paper	E-mail	Paper	E-mail
Letter and Survey	Paper	E-mail	E-mail	E-mail
Thank you / reminder	Paper	E-mail*	E-mail*	Paper
Replacement Survey	Paper	E-mail	E-mail	E-mail

*The E-mail reminders included another questionnaire.

Source: Schaefer & Dillman; 1998:383

The four contact methods were used in order to ascertain not just comparative response rates between postal and e-mail surveys but also to determine the effectiveness of a 'multi-mode' contact strategy.

The results revealed very little difference between e-mail and postal mail response rates (58 and 57.5 percent respectively). Interestingly, the e-mail responses were slightly more complete, and the open-ended questions were filled in with greater detail. In addition to this, the e-mail responses were returned much faster, with 76% sent back within 4 days of the initial mailing (Schaefer & Dillman; 1998:390).

The authors recognise that "coverage error will continue to be a problem for E-mail surveys, at least into the foreseeable future" (Schaefer & Dillman; 1998:390). Even if people do own an e-mail address, many may not frequently check their inbox. Hence, they strongly advocate the use of a mix-mode strategy. Within a known population, this procedure allows those without e-mail addresses to be included in the survey.

One aspect of the multi-mode method employed in the survey, entailed that one group (group 3) was sent a paper pre-notice with the rest of the contacts made via e-mail. It was found that the paper prenotice proved to be less effective. The authors (ibid: 1998:390) speculate that respondents may not have connected the paper pre-notice with the electronic questionnaire and thus "the paper pre-notice failed to serve its purpose of familiarisation". In this event, the questionnaire may be perceived to have been unsolicited and as a result more likely to be discarded. A number of researchers have found that unsolicited e-mails are against 'netiquette' and are likely to be viewed in an unfavourable light, as Witmer *et al* discovered. Shortly after e-mailing their questionnaires, they received several, often very heated e-mails, complaining about the unsolicited questionnaire. In light of this, the procedure was altered to allow for the initial delivery of a covering e-mail, explaining the survey. Respondents were given the option of declining to partake (and thus not sent the questionnaire) by replying with only 'no' in the body of e-mail (Witmer *et al*, 1999:152). This method proved more successful for Witmer *et al* as they did not receive negative responses as previously experienced. However, of the 331 respondents e-mailed the pre-notice, 47 (14.2%) did reply saying that they did not wish to participate.

Alternatively, the researcher could send an introductory e-mail and if respondents are interesting in participating, all they are required to do is simply 'reply' to the e-mail and upon doing so a questionnaire will be sent to them (Comley; 1996:2).

It may be advantageous to allow respondents the flexibility to use e-mail in the way that they are accustomed. Some people are in the habit of printing out their e-mail and attending to it in that manner. Thus, Schaefer & Dillman recommend that participants be offered this option and that a return postal address is included in the e-mail (1998:390). It may also prove to be beneficial in achieving a higher response rate, to send a duplicate copy of the questionnaire with each reminder (as opposed to a thank you / reminder postcard). The nature of e-mail is such that it allows this to be done with limited cost, in terms of time and money.

As with mail surveys, the use of incentives are beneficial in increasing response-rates. Witmer *et al*/ suggest that the inclusion of incentives in Internet surveys is even more crucial, as the technology facilitates dismissal of the participation request with the click of a button (1999:156). As an incentive to complete his survey, Comley offered respondents £2 as well as entrance into a prize draw for a £400 wide screen television. The option to donate the £2 to charity was also offered, which proved to be a popular choice (1996:3).

Non-response in web-based surveys

The issue of non-response is more difficult in terms of a web-based survey. The means of obtaining respondents for this type of survey are similar to that of quota and convenience sampling (discussed earlier in the chapter). Consequently, the

researcher has no real idea of the characteristics belonging to those who chose not to complete the questionnaire. Solicitation of potential participants needs to be undertaken via the use of advertising, banners, notices etc. These methods have proved to be reasonably effective in the case of Survey2000. Further, with the use of existing data, the authors were able to identify deviations from the general population and to draw up weights accordingly. This method seems to be one potential means of ensuring that the data is valid and reliable.

Another way of investigating non-response in a web-based survey is to monitor the number of people who access the questionnaire but, for some reason or other, fail to complete all questions and click on the send icon. Spink *et al* (1999) undertook a survey of EXCITE (a major WWW search engine) users. They placed an interactive survey form on the EXCITE homepage for a period of five days. During this time, approximately 3 729 visitors were logged, 480 of which clicked on the "send survey" icon (Spink *et al*, 1999:3). Thus, only 12.9% of the potential respondents chose to complete the survey, which poses interesting questions as to the characteristics of both those who did and those who declined to partake in the survey. It would be interesting if a mechanism to establish the general characteristics of those not-partaking could be developed. This would bring far greater validity and reliability to such studies.

All in all, the question of sampling and non-response for we-based surveys remains problematic. Some innovations have been made; for example the comparison with existing data and subsequent weighting in the Survey2000 study. However, many gaps remain. Kaye & Johnson (1999:334) state, "traditional sampling methods need

to be reformulated and new ones constructed to accommodate the Web's unique features before it can be considered a viable survey instrument". E-mail surveys measure up better, especially when employing this method when studying an organisation and so sampling become far less problematic. However, other studies using e-mail appear to be more troublesome. Kaye and Johnson (1999:334) argue that an online, e-mail 'yellow-pages' of web users would be beneficial in aiding valid online sampling.

Section two: Questionnaire design

The composition, structure and presentation of questionnaires have long since been identified as another significant source of error inherent to surveys and in particular to postal surveys. Respondents, in postal surveys, are required to complete the questionnaires without any assistance and as a result, it is essential that all instructions are precise and easy to understand. Many debates exist regarding the optimum way to maximise response by means of the questionnaire design. The consideration that the researcher should give to the content and visual appearance is of vital importance. This includes the layout, graphics and colour as well as the length and content of the questions posed to the respondents. For example, Dillman has advocated, since the late 1970's, the use of the Total Design Method, which involves consideration of all of the above points in the construction of the questionnaire (1991:230). The wording used to ask the questions is very important. Standard errors such as ambiguous, leading or double-barrelled questions must be avoided. It is also important that respondents do not feel that any moral judgements will be made. It is easy for lists and sections to appear as if morally right and wrong answers exist.

Finally, it has been shown that shorter questionnaires prove more successful in terms of completion and response rates¹³.

The principles behind questionnaire design remain very similar when applied to Internet surveys; adequate consideration must be given to the details listed above. However, the development of an Internet questionnaire does provide a number of additional challenges, which are discussed below.

Questionnaire design in e-mail surveys

The design of a questionnaire for an e-mail survey is another very difficult task facing e-mail surveys. The questionnaire format can be radically changed with the use of different e-mail programs. For instance, the format of an e-mail opened in Pegasus may look different to the same e-mail opened in a Hotmail account. Schaefer and Dillman comment, "the difficulties of setting up a format that will appear the same on all users' screens are substantial" (1998:392). Further, the size of the questionnaire may also prove to be problematic in an e-mail survey. Some e-mail users may have restrictions on the e-mail length/size (text buffer), causing problems for questionnaires longer than 10 - 15 questions (Bauman *et al*; 1998:4). Thus it is imperative that these considerations are kept to the forefront when developing the e-mail questionnaire. The design process is made easier for the researcher when the survey is conducted within a closed group (i.e. organisation) who all use the same e-mail system. Otherwise, at this stage, for the best results e-mail questionnaire design and size needs be kept as simple and small as possible.

¹³ Several texts provide detailed information in questionnaire design. For example, see de Vaus, 1996; Newell (in Gilbert) 1993 or Rea & Parker, 1992

Questionnaire design in web-based surveys

The most obvious difficulty, which faces a researcher undertaking a web-based survey, is the technical complications associated with placing the questionnaire on the WWW. Not just in the mode of writing a HTML form but also whilst designing the web-form, the researcher must be aware of the changes that will occur to the questionnaire format depending on which browser is used to view the form (i.e. Netscape Communicator vs. Explorer). Fortunately, it is feasible to overcome these hurdles without too much difficulty, as the changes that arise are not as dramatic as in e-mail program, an issue discussed above.

A number of texts exist that provide detailed guidelines on how develop (using HTML and/or JavaScript) and manage web-based questionnaires. For example Baron & Siepmann (2000) offer a step-by-step guide in creating and using web questionnaires. Alternatively, assistance can be found on the WWW itself. Several sample questionnaires have been placed on the WWW, which are helpful to novice web-researchers in providing illustrations of what can be done. The department of Psychology at the University of Plymouth have made available both their tutorial "how to put questionnaires on the Internet"¹⁴ as well as an example of a web-based questionnaire compiled as an illustration to the tutorial¹⁵.

However, the creation of web-forms can be rather tricky and may involve valuable research time. Obviously it is possible to commission a web-expert, however this option is costly and it is likely that most project-budgets would not accommodate this.

¹⁴ <http://salmon.psy.plym.ac.uk/mscprm/form.htm>

¹⁵ <http://salmon.psy.plym.ac.uk/health.html>

A more feasible option is to make use of existing software, which automatically creates the necessary HTML (hypertext mark-up language) or CGI (common Gateway Interface) language. Carver *et al* review two such programs, The WWW Survey Assistant¹⁶ and Senecio's On-line Survey Products¹⁷. They comment that such tools are valuable, allowing many more social researchers the opportunity to conduct web-based surveys (Carver *et al*, 1998:4.2). The Senecio package is licensed and is subject to licensing costs, which would be difficult for small-time researchers. However, the WWW Survey Assistant is available free of charge to academics and non-commercial use. They offer a tutorial to assist first-time users and also offer users the option of placing their survey on their server.

In terms of questionnaire design, a web-based survey holds many advantages for the researcher. For example, it is possible to develop 'skip patterns'. This disables the presentation of questions that are inappropriate as a result of a particular response. This method was successfully used in Survey2000. Witte *et al* remark that one of the major benefits of this is that it shortens survey time (2000:184). The researcher is also able to specify questions that cannot be left blank by the respondent, reducing the number of discarded questionnaires as a result of incomplete data. The WWW also lends itself to the creation of interesting and innovative questionnaires. With relative ease, questionnaires can be colourful, attractive and even interactive - all of which are useful in keeping the respondent interested. A word of caution - a complex, colourful questionnaire with numerous graphical images will increase download time for the respondent who, as a result of time and/or money factors, may decide to

¹⁶ <http://or.psychology.dal.ca/~wcs/hidden/home.html>

¹⁷ <http://www.senecio.com>

abandon the questionnaire before completion (Carver *et al*, 1998:5.1). Finally, another advantage that web-based surveys have is that the respondent cannot alter the format of the questionnaire (Witmer *et al*, 1999:156).

On the basis of their online survey of political information and the web, Kaye & Johnson (1999:331) offer a number of recommendations to consider during the web survey design process:

- The survey should be as short as possible for quick completion and to minimise excessive scrolling.
- Simple designs with sparse use of graphics save downloading time.
- Drop-down boxes save space and clutter by avoiding repeating responses.
- Instructions should be clearly stated.
- Responding to questions should be easy and intuitive.
- Pre-tests should be conducted to measure length of time and ease of completion.
- A check of the survey using different browsers will uncover any browser-based design flaws.

In summary, it is clear that web-based questionnaires hold technical and design advantages over their e-mail counterparts. Until such time that e-mails received are in the same format as they were sent, e-mail surveys will remain at a disadvantage. An option to overcome this is to send an e-mail (similar to the introductory e-mails discussed in the non-response section) to your sample list, which includes the URL (a hyperlink) to the questionnaire. This allows potential respondents to access the questionnaire via a click of their mouse (Kaye & Johnson; 1999:325). Comley reports that this method is not only efficient in terms of response time, but given the choice, most respondents selected this method as opposed to replying by e-mail (1996:7).

Concluding comments

This chapter has highlighted the considerable methodological obstacles present in Internet-based data collection. Substantial difficulties arise in the compilation of a sample that is considered to adequately represent the population to be studied. This difficulty is particularly problematic in web-based surveys. The issue of non-response is equally challenging for both e-mail and web surveys. Nevertheless, it has been shown that a number of techniques have been developed in order to limit this source of error. Finally, we have seen that the technical difficulties in terms of the questionnaire design are substantial for e-mail surveys and perhaps present the most complex hurdle for this method to overcome.

In addition to these limitations, a few more exist that could hamper the research. It is clear that questionnaires returned via e-mail are not anonymous, as the e-mail address readily distinguishes the respondent. As with postal surveys, the pressure to obtain approval (social desirability) is somewhat reduced by the lack of human contact. However, if the research is sensitive in nature (e.g. drug use), the respondent may not be fully honest in his or her answers, or even decline to participate in the survey, as a direct consequence of being identifiable. The possibility of returning questionnaires via an anonymous server exists. However, this relies on the participants' knowledge of the Internet. In the case of e-mail surveys, researchers are able to offer respondents assurances of confidentiality, which may suffice in most instances (Sheehan & Hoy; 1999:5).

Web-based surveys are not exempt from the problems of anonymity and confidentiality. A respondent's identity could potentially be discovered via the IP

(Internet protocol) address of the computer that they used, although, this is not possible if a public computer (such as those an Internet café) is used. Researchers are able to take steps to ensure the anonymity of participants, as Comley illustrates in his study of drug dealers. As normal, respondents completed the questionnaire online. However, as soon as the send button was pressed, the data was placed in a database that did not store any information on its origins, thus, allowing respondents complete anonymity (1996:6.5).

A further hazard for Internet researchers is the possibility of duplicate responses (multiple copies of a questionnaire completed by a single respondent). This is relatively straightforward to monitor with e-mail, although, less so with the increased use of anonymous returns. However, more problems arise with web-based surveys. The monitoring of respondent IP numbers may act as some sort of control, though it should be noted, this is by no means a foolproof method. The IP address was added to the top of each completed questionnaire in Kaye & Johnson's (1999:333) study, which enabled them to keep track and discard duplicate responses. The authors comment that, in the majority of cases, duplicates were sent within a few minutes of each other, perhaps indicating that respondents were unsure whether their posting had been successful.

Notwithstanding the above limitations, Internet surveys possess a number of advantages and hold immense potential for social research. Possibility the biggest advantage of this method of research is that the data is returned in computer format. It is possible to design a questionnaire with the coding in-place, thus eliminating the

costs, both monetary and time, of data capturing. Further, data-entry error is also kept to a minimum.

Swift response rates are reported in most of the existing studies using the Internet as a data-capturing medium (Sheehan & Hoy, 1999; Yun & Trumbo, 2000). Swodoba et al (1997:252) report that 90% of the total participants responded within 4 days of the initial e-mailing of the questionnaire. Potentially, speed of responses such as this, could result in a substantially shorter time frame for the entire research process. It has also been reported that the quality of the response received from Internet surveys is superior to that of their traditional counterparts. Most researchers assert that the responses to open-ended questions are more detailed and hold greater self-disclosure (Yun & Trumbo; 2000:6). Finally, the costs of undertaking an Internet survey may prove to be substantially less than compared to traditional methods. However, although many researchers have speculated that this is the case, it has not been undisputedly established.

It is apparent that while the Internet demonstrates immense potential for future survey research, there are still a number of methodological hurdles that first need to be overcome. The need for research to address the various issues is apparent. In the meantime, clarity in the phases of research during the writing up of the study is important in facilitating confidence in the study.

We began this chapter by considering how long-established sampling techniques (including non-response) of survey research carries over when applied to e-mail and web-based surveys. We found that many of the problems faced by the former (e.g.

ensuring representativeness) are to be found in the latter. We have looked also at the problems facing questionnaire design in both, where we found where web-based surveys have a distinct advantage over e-mail questionnaires.

So far, we have been concerned with (quantitative) survey research, as opposed to research of a qualitative nature. Traditionally, the methods associated with survey research have been regarded as more rigorous and less problematic than those associated with the latter. I now turn to the advantages and limitations of the established ethnographic techniques when they are applied in Internet research.

Chapter Three

INTERNET ETHNOGRAPHIC RESEARCH

Introduction

"Ethnography" is also known as observational research, case study or naturalistic research. "Sociologists usually use this method when they are especially interested in understanding a particular organisation or substantive problem rather than demonstrating relations between abstractly defined variables" (Becker; 1979:312). The underlying logic of a study of this type dictates that a researcher seeks an in-depth account of the social group or phenomenon in question. According to Singleton, naturalistic observation produces a "social picture describing people in their own terms" (1993:345). Very often, ethnographic methods are employed when little or no information is available on the study topic and the researcher undertakes 'path-breaking' research (Fielding; 1993:155). Alternatively, ethnographic research can be used to make the known 'problematic', in an attempt to make the familiar, unfamiliar and by doing so gain better understanding of the phenomenon (Hine; 1998:4). It is worth noting here that an extensive debate surrounds ethnography, and indeed any research of this nature, which questions the overall validity of research with such an interpretivist nature. The debate oscillates between two extremes; at the one extreme those who fully advocate the method-- e.g. Blumer, and at the other those who deem it 'unscientific' -- e.g. Lundberg (Hammersley; 1989:chp.4).

As stated in Chapter One, the number of people regularly connecting to the Internet globally exceeds 400 million. With this growth, there has been a rise in ethnographic research opportunities/problems for social scientists. The existence and development

of the so-called 'virtual communities' have been a concern to researchers for a number of years. Members of such virtual communities claim to feel a deep sense of belonging and are often very protective of their group – a phenomenon that is already receiving attention by social researchers. It is also argued that the network opens up previously inaccessible areas for study. For example, research involving people living in rural and remote areas, members of various groups – often those that are unusual or deviant - and cross-national studies are perhaps more feasible with the use of the Internet (Lee; 2000:117/8). Another, oft weighty, consideration in favour of Internet ethnographic research is that costs can be kept to a minimum. The reduced cost is even more likely than with Internet survey research.

This chapter, divided into two main sections, examines the use of ethnographic methods in Internet studies; i.e. the Internet as the main means of gathering ethnographical data. The first section is devoted to a consideration of the actual techniques used for cyber ethnography, by examining a number of studies employing this method, involving the use of virtual participant observation and online interviewing. I will then discuss a few central aspects of cyber ethnography; the role of the researcher, reflexivity and the function of interpretation. The second section in this chapter deals with a number of ethical issues encountered in cyber ethnography. This includes the problem of informed consent, the blurred distinction between public and private spheres, covert research, and the possible benefits and harms of ethnographic research.

Section one: Cyber ethnographic methods

By using of ethnographic methods, the researcher aims to gain an in-depth understanding of the problem at hand, e.g., the rules that govern a society, the language used by a particular group, relationships between the social actors etc. Although not highly generalisable, the rich data obtained is useful in understanding the intricacies and binding characteristics prevalent in the community. In conventional research, observation (participant and non-participant) is often the main means of gathering this type of data. However, in-depth interviewing of informants is another method, often used in conjunction with observation. The same basic principles of ethnography apply to cyber-research.

Traditionally in ethnographic research, a researcher often depends on cues given by elements such as physical appearance, facial expression, and voice intonation of the informant. This raises questions as to how researchers undertake 'observation' on the Internet, where visual and sound cues are absent and observation is limited to the reading of texts. I would argue that to fully understand an online community, it is fundamental for the researcher to experience exactly what the members of that community experience - and that is text-based communication taking place in a 'virtual' society.

This lack of paralinguistic cues is a point that Rice-Lively comments on in her study of a networking class. "My ethnographic research techniques primarily were dependent on the electronic communication shared in the community newsgroups" (1994:3). Despite the absence of visual and sound cues, Rice-Lively reports that the limitations do not place any question marks on the value of the research. Quite simply, the

research methodologies must be adapted to suit the new research area. Thomsen *et al* (1998:7) also argue that, with the development of methodological adaptations to cater for the absence of paralinguistic cues, "careful analysis of the online discourse or text can lead the ethnographer to an appropriately 'thick description' of the online community". This issue, as well as the practicalities of gathering ethnographic data via the Internet, will be investigated in the following section.

Virtual observation

According to Goffman, participant observation involves the researcher becoming a part of what he or she is observing. This is done in order to familiarise and make sense of the community under scrutiny. He states:

"any groups of persons - prisoners, primitives, pilots or patients - develop a life of their own that becomes meaningful, reasonable and normal once you get close to it, and... a good way to learn about any of these worlds is to submit oneself in the company of the members to the daily round of petty contingencies to which they are subject" (Goffman; 1961:ix-x cited in Fielding; 1996:156).

As we shall see, these concepts remain constant for participant observation undertaken on the Internet.

Smyres (2000) undertook a cyber ethnographic study using participant observation to investigate young (adolescent) girl's perceptions of their bodies. Smyres (2000:2) writes, "I observed participants in what appears to be a safe space where I would be unobtrusive, a website devoted to teenaged girls and adolescent struggles". Data was collected over a six-week period and a total of sixty hours exploring all areas of the

website. Through these steps, she claims to have; "gained extensive knowledge of the sense of community the staff members of gURL attempted to foster and how adolescent girls interact within that context" (2000:6). All aspects of the website were examined by following links within the site. Sixty articles by website reviewers and newspaper articles were also taken into consideration. Finally, and most importantly, the discussions on two bulletin boards ('Do guys judge girls' and 'Exercise') were analysed. The researcher remained unobtrusive, i.e. her presence was not made known (the ethics covert such as this will be discussed later) and so did not actively participate in any of the discussions, nor did she initiate contact with any participant. This would appear to have been a form non-participant observation. However, Smyres claims that through the nature of such a website, she was a participant as well as an observer. She argues (2000:7) that the role of participant is gained through site registration with a username, password and e-mail address. Thus highlighting an interesting unique feature of the Internet, which will be further addressed later in this section.

Another example of participant observation in an online community is found in Pleace *et al/s* (2000) study of a support group for problem drinkers. Data was collected over a period of five months during which time the chat-room was visited 54 times, ranging between one and two hours in length at a time. Over 300 conversations between members were recorded and later placed into broad categories and analysed. Like Smyres' study above, the authors of this study did not seek permission, nor reveal their real identity. The authors report (Pleace *et al*; 2000:2.4) that their decision not to declare themselves was partly made in view of the sensitive nature of the community and thus concerns of participant reactivity to the researcher's

presence were real. If questioned by other participants, the researcher (Pleace *et al*; 2000:2.5) stated that "they were 'curious' about the room, but did not claim to be a problem drinker and remained silent during those exchanges in which individual sought advice about their problem drinking". The ethics of such a study, undertaken covertly, will be examined in section two of this chapter.

Participant observation can also be employed in the study of online discussion groups. Fox and Roberts undertook an investigation into the social order of a medical list for GPs, whereby the researchers became members of the *gp-uk* e-mail discussion group. The authors focus on "the reasons given for participation in the discussion list, informal rules and norms, and the social construction of members' cyber-identities" (Fox & Roberts; 1999:652). Data was gathered over the period of a year, where both researchers joined the list: one, Chris Roberts, joined as an active (participating) member of the group, openly declaring his status as a researcher after gaining permission from the list owner; the other researcher, Nick Fox, joined the group but only in the capacity of a 'lurker', a non-participating member of the discussion group. All messages to the list during the data-collection period were read and saved to the principal database. The authors report that other data used in the analysis included (1999:652):

- private mail generated by CR's participation in the list
- discussion between the authors of the paper and with other researchers of virtual communities and
- searches of the *gp-uk* archived messages.

The study of online communities is not always limited to virtual contact with participants. Rice-Lively (1994:4), who conducted an ethnographic study of a

networked learning community of an Internet based graduate level course, followed four steps during her data collection period:

- observing students and the instructor during eighteen hours of personal sessions;
- reading each posting to the class newsgroups and saving messages that she believed illustrated cultural patterns;
- tape-recording study informant's interviews. This involved a minimum of fifteen hours of prearranged interviews sessions with selected class members, and a number of short interviews with students who volunteered information during the face-to-face classes and through e-mail;
- writing observations and discoveries in a field journal at the end of a day or muttering a new observation or a question into a tape recorder when travelling to and from work.

As one can see, this study used a mixed-mode approach, with a certain degree of face-to-face observations/interviewing combined with the analysis of the online communication. This combination of virtual and 'real life' contact is also used for interviewing participants, (e.g. Jacobson; 1999) which will be investigated later in this section.

Other methods of online ethnography are not as evidently similar to the 'real world' methodology. Leaning (1998) undertook an examination of identity and meaning in a chat-room. Much of the data was recorded by "simply open(ing) a chat-room window on my p.c. left it open for a number of hours and then copied the resulting conversation from the chat window to a word processor" (Leaning; 1998:5), a method which he refers to as 'covert non-participant observation'. "Covert" obviously as he declined to inform those participating of his actions, rather it is the 'non-participant observation' that I find somewhat problematic. In Smyres study, discussed above, she

claims that by virtue of registering as a member of the community, following real-time discussion and examining all aspects of the website, she became a participating member of the community. That I would agree with. Leaning, on the other hand, simply chose a number of Geo Cities public chat-rooms and recorded the going-ons in the manner previously described. Further, no time element was captured thus leaving the author with little idea of the speed of responses. Another study using similar data collection techniques was undertaken by Paolillo (1999), who carried out an exploratory investigation of social networks and language on an IRC chat-room (#india). The author reports "interaction on #india was recorded for a complete twenty-four hour period by connecting with an IRC client program and capturing the entire session to a log file" (1999:7). I would argue that this method borders more on a form of content/documentary analysis rather than a form of 'observation', by virtue of the fact that they observed/read the data after it had occurred.

Thomsen *et al* (1998) also take this view. They argue that 'prolonged engagement and persistent observation' are essential features of ethnography and this does not change in online ethnography. They state (ibid; 1998:10) "the researcher needs to be willing to commit as much time to online ethnography as he or she would to a more traditional on-site study....One cannot sample a community's experience from the records, even if you have access to everything that has ever been said". Thus, according to Thomen's criteria, the above studies by Leaning and Paolillo cannot be thought of as ethnographic studies but more as a form of textual analysis.

Online interviewing

Another key feature of ethnography is the interviewing of informants - be it one-to-one or as a group. It has been suggested that online interviewing (as opposed to 'real world', face-to-face interviewing) removes a number of potential bias from the data. Selwyn & Robson (1998:2) maintain that concepts of race, gender, age and sexuality are not applicable in online interviewing. Another obvious advantage for the researcher is that virtual interviews eliminate the need for costly and potentially error-inducing transcription of tapes into computer format. Online interviews can take the form of asynchronous e-mail interviews, which often take place over a certain period of time. This is useful in many ways. Neither geographic location nor time-constraints feature in this method (Selwyn & Robson; 1998:3). Both the interviewer and interviewee are free to respond at their leisure and perhaps this gives the participant extra time to consider his or her responses¹⁸. Another form of online interviewing involves the use of the WWW for 'real-time' (synchronous) interviews with the researcher and participant¹⁹.

Jacobson (1999) undertook a study on the formation of impressions during online interaction. Mentioned earlier, this study used both on and offline participants. The two groups of informants comprised of, one group of fifteen online respondents and another group of twenty-three (real world) students, who were participating in a course on 'social relations in cyberspace'. The online informants were volunteers who

¹⁸ Gaiser (1997:139) found this to be the case when using an e-mail discussion group to conduct focus groups over the period of a week.

¹⁹ For example, Ward (1999:5.3) used the interactive bulletin board system (BBS) to undertake semi-structured interviews of selected participants in the community where she was a participant observer.

responded to the author's request for participants, which read (Jacobson; 1999:8):

"For research on the ways in which MOOers (and other participants in CMC) form impressions of one another, I'm interested in talking with people about their experiences in meetings MOOers offline. I'm particularly interested in the 'fit' (or lack thereof) between expectations based on MOO interactions and 'rl' (real life or offline) experiences. Are people what you expected them to be? If not, why not? In what ways do their online and offline characteristics differ? If you are interested and willing to take time to talk, please moomail me."

Each respondent was interviewed, online, about his or her experiences. Interviews ranged between one and two hours and conversations were logged (recorded) with the participant's permission. In a few cases, additional information was sent via 'moo-mail' - a form of e-mail internal to a virtual community (ibid; 1999:8).

The offline (face-to-face) respondents were asked to form impressions of a total of twenty-seven MOOers, selected from a list of thirty-five, each of whom had placed photographs online. After they had documented their impressions, the participants were shown the photographs, interviewed on their impressions and whether these impressions corresponded with the visual appearance of the MOOer. Unfortunately, the author provides little information on how the online interviews were conducted, problems encountered or indeed advantages of this method. However, Jacobson (1999:9) did say that he "took the informants' offline meetings as a given", choosing not to enquire into the details of where or why the meetings had taken place. This raises an interesting point concerning the element of trust involved in online research. In the case of anonymous participation in an online study, the researcher can only

trust (and hope!) that the participant is telling the truth. Verification of most information given online is difficult, especially the issue of whether or not that the participant is who she said she is; an issue which we shall deal with when discussing informed consent later in this chapter.

Chen & Hinton (1999:1.2) investigate in depth the use of standard WWW browsing software²⁰ (i.e. Netscape) to introduce a method for real-time interviewing. According to the authors, the essential characteristics of WWW interviewing are (1999:2.1):

- it is conducted in real-time, attempting to capture the spontaneity of traditional interviews;
- it uses existing software in a novel manner, rather than requiring the acquisition or development of new programs;
- it places the technical burden on the researcher, reducing effort required to participate on the interviewee; and,
- the method is essentially an application of an existing technique, rather than a wholly new research approach.

A number of interesting features emerge from Chen and Hinton's pilot study. The authors (1999:12.4) report that, by using the new technology, the interviewer enjoys less control over the process, with respondents feeling more in control and often less nervous. Participants, however, also reported a pressure to respond faster than they normally would (as a result of the visual absence of the interviewer) as well as the inclination to type before they had fully thought-out their response to the question (1999:12.7). Chen and Hinton conclude that the method brings four major advantages: real-time interviews regardless of distance and cost, full transcript of

²⁰ In this, 'realtime' interviews are conducted by means of a webpage, which is used as an interviewing 'screen' between the interviewer and the interviewee. The interview is logged (recorded) thus providing an instant transcript. (Chen & Hinton;1999: 4.1)

interviews is available immediately, distancing of interview from interviewer and the use of computer software to assist in the process of collecting and subsequent analysis of data. They also concede that a number of limitations exist in the use of this approach. For example, the lack of paralinguistic cues and observation proved to be problematic during the interview. Questions were also raised as to whether this method lead to data reduction (less information provided by informant) and further bias introduced by sampling limitations is problematic (1999:14.2). For instance, clearly such interviewing could only take place where all participants are computer and WWW literate.

In addition to conducting researcher to (one) informant interviews, a researcher may also undertake interviews with a number of informants simultaneously (more commonly known as focus groups). Stewart *et al* (1998) not only undertook online focus groups but they did so in order to draw comparisons between two geographically remote countries, Australia and China. The authors of this study investigated young women's perceptions of health risk (in particular, to tobacco use). The focus groups were conducted online after introductory face-to-face focus groups were conducted at each of the four venues, where an onsite facilitator oversaw the subsequent link up and duration of the online focus group. A total of sixteen young women between the ages of eight and thirteen participated. They were recruited by 'snowball' sampling methods (Stewart *et al*, 1998:4).

The researchers used Firstclass Conferencing software that allows synchronous communication. The chat-rooms were configured to allow for private access only. The focus groups were expected to last an hour, although the actual duration was closer

to two hours. Interpreters / translators were used in Beijing and thus delays in the receipt of data were experienced. Again, understandably, the participants who were faster typists had more to say than their slower counterparts. Indeed, Steward *et al* concluded their paper by highlighting their view that the typing abilities of each participant are central in the attainment of a successful interactive online discussion.

Another problem that the focus group facilitator experienced in Steward's study, was a lack of control over the discussions. Traditionally, the facilitator is able to guide the discussion and, where necessary, to get everyone back on-topic before the conversation has strayed too far. In their study, however, Steward *et al* (1998:9) comment "the type of data generated seems in part a result of the facilitator's inability to exert sufficient direction and/or control over the discussion, as would have been possible with face-to-face focus groups".

Aside from the Firstclass Conferencing software, a number of other software packages are available to facilitate both online interviewing and focus groups - many of which are user-friendly. O' Connor and Madge (for their study on CyberParents, 2001) made use of Hotline Connect software²¹, which, like the Firstclass Conferencing software, enabled real-time, group or one-to-one chat to occur at a specific address that was determined by the researcher. An important feature of this software was its accessibility and ease of use. The researchers were not present to deal with technical difficulties that arose, although technical (installation) assistance was available, via telephone or e-mail prior, to the interviews.

²¹ <http://www.bigredh.com> - the Hotline Connect software website.

O' Connor and Madge also experienced problems similar to those reported earlier. The lack of verbal cues to encourage participants to continue or elaborate proved to be particularly problematic. The authors feel it is essential that alternatives must be found to overcome this hindrance. O' Connor and Madge also note that in this type of study, researchers are reliant on the computer literacy and familiarity, in order to achieve maximum input from all participants. However, they did find that this method possesses a few clear advantages. These include factors such as the absence of the researcher, location of interview and travel time that become irrelevant when using the Internet. They also comment upon the ease of gathering data into a Word document for analysis.

So far, we have looked at the methods, advantages and disadvantages present in cyber ethnography. We shall now turn our attention to the role of the researcher in this context.

The role of the researcher

All social research involves a researcher who is inextricably part of the social world he or she wishes to study. However, more uniquely for ethnographic studies, the researcher is pivotal in the actual data collection process; "the researcher becomes the research instrument" (Rice-Lively; 1994:2). Thus it follows that at every stage, the researcher may potentially influence or even bias the research. For example, the researcher may have a preconceived opinion on a particular subject that could prevent him or her from fully understanding the position of the people/community in question.

In order to address the above concerns, many commentators advocate the use of reflexivity, which openly acknowledges that the researcher is part of the very world that he or she wishes to study. That is to say, they need to admit that their stocks of knowledge and life experiences have been derived from this world. Reflexivity, to Hammersley and Atkinson, means

“a rejection of the idea that social research is or can be carried out in some autonomous realm that is insulated from the wider society and from the particular biography of the researcher, in such a way that its finding can be unaffected by social processes and personal characteristics”

(1995:16).

Acknowledging the position of the researcher, however, is not to imply that valid social research cannot be undertaken. Indeed, it is often claimed that this recognition of the researcher's position is fundamental in aiding the collection of representative research. Thus throughout the research process, it is essential that researcher reflexivity is central.

Within cyber-ethnography, reflexivity has proven to be a core feature. Ward (1999:1.13) explains:

'...with cyber-ethnography, expectations and traditional definitions of the situation are dropped, as the researcher has very little control over the conversation. Participants remain unknown to the researcher and this adds to the balancing of power between the researcher and researched. The participants have no reason to trust the researcher,

and are therefore in a stronger position to ask questions and challenge the assumptions of the researcher. It is this potential that the participants have for challenging that makes the research process reflexive".

Fox & Roberts also argue that reflexivity increases in cyber ethnography. They suggest that it (reflexivity) is heightened by the textual nature of online communication; "the ethnography and the community can no longer be separated, and reflexivity is also total" (Fox & Roberts; 1999:651). This position suggests that the idea of a scientifically neutral perspective in social investigation is enhanced when the unavoidable involvement of the researcher is openly acknowledged and addressed.

Interpretation

By the term 'interpretation', I am not referring to the process of interpretation that takes place in the data analysis phase of social research. Although an important issue, here I am interested in the question of interpretation during the data collection stage. It has been argued that meaning only comes from 'acts of interpretation' (Steedman; 1991:54). In everyday life, we interpret what we see from the perspective of our existing body of knowledge. Social scientists are not exempt from this. In many instances, they enter unknown situations and through the process of observation and participation, they attempt to make sense of what they are seeing, not from their existing body of knowledge but the perspective of those being studied. This is achieved by prolonged exposure to the research community. If successful in their fieldwork, they will hopefully come to understand the actions/events in much the

same way as those whom they have studied. However, this is where the fundamental problem lies. How does the researcher know he or she has interpreted a word, action or event in the intended manner? This is a point that O'Connell Davidson & Layder (1994:25) note, "you can never know with absolute certainty that another person understands what you say in *exactly* the same way that you understand it...language, concepts and beliefs affect our *perception* of social reality".

Of course, at the end of the day, the researcher's interpretations of the community he or she has been studying may differ from the interpretation of the same event by those researched, something Roy Wallis experienced in his research on Scientologists. In an attempt to accommodate their views (and to prevent libel action) Wallis (1977:165) provided key members with a copy of his final manuscript so that they could give their interpretation of the account of the shared experiences.

In a similar vein, Hammersley and Atkinson (1995:131) argue, "all accounts must be interpreted in terms of the context in which they were produced". They go on to suggest a good way for a researcher to verify his or her interpretation is to cross check with informants. This provides perhaps the best safety net against mistaken interpretations in the researcher's observations. For instance, Whyte's intimate relationship with the key informant, Doc, with whom he discussed many of his observations and ideas, greatly increased the validity and reliability of his interpretations (Whyte; 1994:73). Another example is Laud Humphreys' (1970:36) "intensive dozen", who performed a cross-referencing role in his study on impersonal sex in public restrooms. It is through these steps that conventional research has attempted to overcome the problems surrounding interpretation.

In the case of cyber ethnography, Ward (1999:4.5) found that, owing to the ease of access to the research field and hence the open nature of such research, the interpretation process "remains open to constant re-negotiation by both the researcher and the researched". It could be argued that this continual referral to the participants - and clarification of issues troublesome to the researcher - further enhances the validity and reliability of cyber research. Thus when using the Internet for data-collection, there can be on-going, open-ended checks of the participant's interpretation of relevant events.

Before proceeding to the following section addressing the ethical considerations in such research, I shall first summarise some of the important points we have noted. This section has illustrated how researchers are approaching cyber ethnographic studies. An issue that we opened with, the lack of paralinguistic cues has proven to be a challenge for all researchers. However, some have found their absence more problematic than others. It would appear that this is the case in online interviewing, where all of the researchers in the studies examined, commented negatively on the lack of visual and sound cues. They also feel that the interview is less free, as a result of the turn-taking for question and answer, which breaks the discussion flow. Interestingly though, Chen and Hinton comment, on the loss of paralinguistic cues; "while many social research texts stress the importance of observations and non-verbal cues, few provide the information required to fully interpret them' (1999:12.3).

We have seen that various methods (some old, some adapted and some new) exist to make cyber ethnographic research possible. However, the ethical concerns of any

ethnographic study have proven to be exceptionally difficult. It is to these concerns that we now turn.

Section two: Ethical considerations

Within the realms of traditional social research, ethical concerns are not altogether resolved. However, over the past fifty years, the effects and potential effects of research have been given lengthy consideration in contrast to the days of early social researchers such as Malinowski and Mead. It is possible to look at a number of the early studies and be horrified at the apparent disregard of the researcher for the people who are the subjects of such research. In many cases, issues such as distinctions between the public and private spheres of people's lives, the seeking of informed consent, and the long-term (potentially negative) effects of a study were not considered.

Today, in many instances, a proposed study is often subjected to many hours of scrutiny and debate in order to ensure that the study is ethically acceptable, or and that the participants (and even the researcher) do not face undue harm as a result of the proposed research. Further, over the past twenty/thirty years, many institutions and associations have developed so-called 'codes' to assist their researchers.²² However, the regulatory systems in place vary from institution to institution, country to country and they often fluctuate in strictness as well. Countries such as the United States and Sweden, for instance, are more rigorous in following the ethical guidelines,

²² For example:

American Anthropological Association; <http://www.ameranthassn.org/committees/ethics/ethcode.htm>

American Sociological Association: code of ethics, <http://www.asanet.org/ecoderev.htm>

The British Psychological Society: ethical principles for conducting research with human participants
<http://www.unn.ac.uk/academic/ss/psychology/handbook/99/handbook/Pt21.html>

compared to the United Kingdom, where there is minimal formal intervention (Lee; 2000:135).

Despite these codes, it remains that ethical issues in social research are not completely resolved. Indeed the very nature of ethics dictates that there are no standard right or wrong answers. Homan (1991:1) defines ethics as "the science of morality: those who engage in it determine values for the regulation of human behaviour". This may be so, however, these complex matters are engulfed in grey – some darker or lighter than others but rarely, if ever, simply black or white.

Cyber research is not exempt from these problems. Indeed, it could be argued that the biggest challenge facing a cyber-ethnographer are the ethical issues that occur. For example, the adumbrate distinction between public and private spheres is most problematic when studying an online community of some sort. Some commentators argue that users of the Internet have no expectation of privacy, as this is a public forum. In contrast, others maintain that expectations of privacy are legitimate and should be respected. If it is determined that the study is taking place in a private sphere, the problem of seeking (and gaining) informed consent also arises. On the other hand, if it is deemed that the research is taking place in a public sphere, and the researcher chooses not to declare his or her intent, nor seek informed consent of participants, then a number of moral questions hang over the research. The remainder of this chapter focuses on these ethical issues, commencing with a discussion of informed consent in social research.

The problem of informed consent

It is likely that the majority of research takes place with the knowledge of participants. Indeed, the gaining of informed, voluntary consent of participants has long been considered to be an imperative. In recent times, the notion of 'informed consent' was developed during the Nuremberg War trials (1949), which stated:

"The voluntary consent of the human subject is absolutely essential. This means that the person involved should have legal capacity to give consent; should be so situated as to exercise free power of choice, without the intervention of any element of force, fraud, deceit, duress, over-reaching or any other ulterior form of constraint or coercion; and should have sufficient knowledge and comprehension of the elements of the subject matter involved as to enable him to make an understanding and enlightened decision"

(Reprinted in Burgess; 1984:200)

The convention above requires that in any form of research (although it was initially addressing biomedical research) full consent must be obtained, after completely explaining the nature of the research to the potential participant and ensuring their understanding of the explanation. Further, they must be legally able to give this consent (i.e. consent must be sought from legal guardians of minors). Finally, it is the right of the potential participant to refuse consent and therefore not participate in the study.

Although this is clear in theory, several problems arise with this in practice. Firstly, how can one ever be completely sure that another person "fully" understands what is communicated to them? Secondly, very often, especially in the case of ethnographic research, the researcher is unclear herself as to what direction the study will take. Therefore, the best he or she can do at the beginning of the research is to explain the

initial intentions of the study. However, this raises another problem. If, during the course of the field research, a different direction is taken, should the researcher then return to his or her participants and 're-apply' for their informed consent. And if so, at what point is the researcher obliged to re-inform his or her participants?

Finally, the potential effects of full informed consent on the research itself must be considered. If participants are aware what the researcher is looking for/observing, they may well change their behaviour – the "Hawthorne effects" has been well documented. Let me outline them briefly, Elton Mayo undertook an investigation into worker productivity at the Hawthorne works of the Western Electricity Company in Chicago. Regardless of what variables were changed, the productivity of the workers usually increased. Mayo realised the workers were actually responding, not to the changes in their environment, but to the knowledge that they were being observed (Haralambos & Holborn; 1995:813).

A related problem that arises here is that in some situations it is impossible to gain the informed consent of those being observed. As Punch (1986:36) states; "it is not only physically impossible to inform everyone of one's purpose and identity but also damaging to the research itself". He uses the example of observing police officers in the process of breaking up fights, asking if he should shout: "freeze, and whip out my code... and chant out the rights of participants" (1986:36)! These are some of the issues that surround the problem of informed consent in conventional research. The same considerations hold true for cyber research as well.

Most researchers accept that if one is engaged in researching a virtual community (even if in an accepted "public" domain), it is only polite to inform members of that community that they are part of a research project. Others writers go further. Denise Berg argues; "no matter what, I think that even if you are a participant-observer... you need to reveal your role as a researcher" (message posted to Online Research, 19th September 1999). However, others maintain that there are no hard and fast rules that dictate this, and the final decision lies with the researcher. It has also been noted that there is an element of risk in informing an online community of the study. It could be that the group responds well to the researcher and is happy to co-operate. However, there is always the danger that the reverse will occur (George Stricker, message posted to Online Research, 20th September 1999). Researchers need to be aware of the reactive nature of seeking informed consent (Lee; 2000:136). The Hawthorne Effect is still applicable on the Internet and measures need to be taken to overcome the potential changes in behaviour of participants as a result of this. However, the bottom line remains there is a wide spread belief that the informed consent of participants is a requirement in morally acceptable research.

Additional issues have arisen as to the legal aspects of recording online interactions without informed consent. This appears to be as problematic as the ethical ones. Many questions have been raised as to whether informed consent is legally required when recording interaction on the Internet. It appears that in general this rests with the expectation of privacy during online communication. "Obviously where the whole Internet world has access to the message, there can be no expectation of privacy. The message can be accessed, copied, repeated and published without (legal) consequence" (message posed by Lin Collette to Online Research, 18th October 1999).

Although this may be a general rule of thumb, it is important to note that laws differ from country to country. A social researcher undertaking an online study would be wise to investigate the legal position of the country where the research is being conducted before commencing the project.

It has been noted that there are many practical hurdles to be overcome when attempting to obtain informed consent in cyberspace. The first issue that arises asks; "who" is the informed consent required from and further, how is it to be obtained? The nature of the Internet is such that it allows people to enter and leave communities at random. Therefore, the study population is bound to be unstable, making it difficult, if not impossible to obtain informed consent from all participants (Lee; 2000:136). To overcome this problem, Ward (1999), before undertaking participant observation in two online communities, decided to inform participants of her research. This was done by posting messages to the bulletin board systems and e-mail lists of both communities, after which she (ibid. 1999:1.10) maintained "it became the participants responsibility to read the message". Along similar lines, it has been suggested that consent can be sought from the list/chat-room/page owner. However, the question must be asked if this is an adequate way of dealing with the problem? It doesn't necessarily follow that the list owner etc. is in a position to give consent on behalf of the group members (Lee; 2000:136).

Another issue that arises in the online pursuit of informed consent, is the lack of 'proof' the researcher has concerning who his or her participants are, their age (especially important in the case of minors) and that they understand the nature of the research. Natilene Bower asks, "how does one provide such 'proof' when doing

research in a space where participant and researcher identities are always questionable?" She goes on to claim that the notion of informed consent in cyberspace may not be pertinent and suggests the creation of ethical alternatives (message posted to Online Research, 14th October 1999). Although what these alternatives would be remain unclear.

A way around the problem of gaining informed consent is the suggestion that this consent in cyber studies is only necessary if data is collected in a private sphere. The American Association for the Advancement of Science (AAAS), in a report on Online Ethics, asserts researchers do not need to obtain informed consent if undertaking data collection in a public sphere (November 1999). But this assumes we have a clear distinction between those areas that are essentially private and those that are public. Hence we now turn to the problems surrounding public and private distinctions.

Public/Private distinctions

One of the most problematic areas in virtual research is determining what is public and what is considered private on the Internet. The existing ethical guidelines are difficult to apply to the study of computer-mediated communication, since the public and private distinctions are not clear in cyberspace (Lee; 2000:135). Given that most sites are accessible to the (global) public, anyone from anywhere can access a web page, all that is needed is the URL address. More secure sites do exist of course; however, these are always subject to intrusion by hackers. In short, very little, if any, information posted on the Internet is safe from prying eyes. Further, it could be argued that the majority of Internet users are aware of this fact. In the light of this, do we then consider the Internet to be an unavoidably public sphere? As stated

earlier, Pleace et al (2000) in their study on problem drinkers decided not to reveal their true identities to the community. This was partly due to fears of reactivity but also partly because they believed the community to be public. They made this decision on the basis that "this was a public forum and that the exchanges within it were within that context as distinct from the private discourse that would be found in a letter, telephone call or e-mail" (Pleace et al; 2000:2.4).

Cavanagh (1999) argues that those who occupy an area determine whether it is public or private and this is so with the Internet, where the members of a community may establish areas that are considered private. She states; "two areas may be considered to be revealing in the context, the views of the users as expressed through the conceptual and verbal apparatuses of the environments and the actions of the Internet community in the case of 'lurkers'". By this, she is suggesting that the tolerance and acceptance of 'lurkers' by active members, confirms that this is a public sphere. If they were unwelcome, lurking would be considered akin to eavesdropping (Cavanagh; 1999). On the other hand, if we follow the so-called 'netiquette', it is accepted that 'private' areas do exist. For instance, to gain access to certain pages/chat areas etc. membership may be required, therefore creating a community, which can be regarded as a private domain.

The debate concerning what is public and what is private is certainly not limited to the Internet. Social commentators, often using past studies to illustrate their point, have discussed the matter in detail. A now notorious case, which has been debated furiously, is Laud Humphreys' (1970) TeaRoom Trade: Impersonal Sex in Public Restrooms, with which comparisons, to virtual issues of public and private sphere, can

be drawn. This study, which was undertaken in the late 1960's, has since been widely criticised from a moral perspective. Not least of these issues is the debate over public/private spheres. If we first address Humphreys' role in the restrooms, as the title suggests, he conducted the majority of the study in a public sphere. However, is it right to assume that a rest room is a public place? It could be reasonably assumed that the men who performed homosexual acts in the public lavatories were under the impression that their acts were being conducted in privacy, hence closed to outsiders, those not in the community. In other words, the privacy of the acts trumps the public nature of the place where these took place. Although the debate on Humphreys' role as a 'watchqueen' was fierce, the majority of criticism was directed at the second part of the study, where he invaded the privacy of the men's homes.

"I would condone Humphreys' approach in posing as a 'watchqueen'.

But once (he) moved to collecting license plate numbers, tracing names and visiting homes under the guise of a different project, then that represents both a severe invasion of privacy and also a potentially dangerous situation if his data fell into other hands. That is going to far for me"

(Punch; 1986:40-1).

We can see from the discussion above that the main problem with public and private spheres is that "what is public and what is private is rarely clear-cut" (Hammersley and Atkinson; 1995:267). Deciding what constitutes as public and what constitutes as private on the Internet is proving to be equally problematic. Some commentators argue that, for example, messages posted in a list-serve group are technically public. Thus Sherri Conden states "these messages are public in the sense that anyone who

posts a message knows that it could be read by anyone who subscribes to the lists, which is potentially anyone who can read and write" (message posed to Online Research mailing list, 20th September 1999). Jason Rutter takes an opposing position to this stating, "although messages are posted into a public sphere they are not necessarily intended for public consumption" (message posted to Online Research; 21st September 1999). That is to say that although a community member is aware that his or her message is posted in a public domain, the content is only intended for other members of the community and not the public at large.

Let us return to the debate over Humphreys' study – private interactions taking place in a public sphere. It could also be argued that the men knew that their conduct could become public and being observed was one of the risks they took. It can be claimed that information, which is obtained from public places, is "up for grabs and those who write, speak or behave publicly, bargain for the possibility that they will be observed – even by the social researcher" (Homan; 1991:170). Similar arguments can be made with regard to some ostensibly private interactions on the Internet. For example, members of a discussion group may feel that their interactions are private. However, it has been pointed out that the majority of groups explicitly state in their welcoming messages that it is a public forum and members should expect what they say to be treated as such (George Lessard, message posted to Online Research, 23rd September 1999). Obviously the difficult moral issue remains. One way of dealing with it is to assume that all information concerning an individual's character or actions is private and that the burden of proof lies with the researcher to show why this should be regarded as of public interest.

The case of covert research

A social scientist is understood to be employing covert field research when his or her identity and the true nature of his or her work is purposefully kept unknown to the people being studied. This practice, while useful to social scientists, clearly presents numerous ethical dilemmas to the researcher. As Erikson writes;

“the practice of using masks in social research compromises both the people who wear them and the people for whom they are being worn, and in doing so violates the terms of a contract which the sociologist should be ready to honour in his dealings with others”

(cited in Bulmer; 1982:9).

Many critics believe covert methods of data collection are always unacceptable and their arguments focus on the potentially harmful effects it could have on participants in such research. Firstly, the research disregards the principle of informed consent. The participants/subjects are unaware of the study and their participation, and are therefore unaware of any foreseeable outcomes and have no opportunity to remove themselves from this danger. Further, without the knowledge that they are being observed, often a close relationship is built up, especially between the researcher and key informants, and a trust develops. The social scientist can be seen to be betraying that trust as well as invading the personal privacy of his or her informant by gaining access to areas that would not normally be open to them (Homan; 1991:110). It is also argued (Hammersley and Atkinson; 1995:266) that covert methods discriminate against the weak and powerless as these groups do not have the influence or ability/power to defend themselves against the research. In this way, the behaviour

or interests of the group being studied could be damaged with adverse effects occurring, such as unwanted/undesirable public attention.

Many of the issues raised above apply directly to the Internet and certainly, covert studies can be undertaken with ease in cyberspace. If the community is acknowledged as private (not readily open to outsiders) and the researcher decides to proceed without informed consent, he or she must carefully consider if the research should proceed at all, or at the very least, how best to protect the interests of the participants/community being studied. One measure to protect the community would be to conceal the identities of participants and the community as a whole, an issue to be discussed in greater deal in the following section.

Bulmer (1982:217) states; "the use of covert participant observation requires most careful considerations". Issues such as confidentiality and misuse of data, should be examined and addressed before any study is undertaken, and further reviewed and revised throughout the duration of the research. Nevertheless, it cannot be ignored that covert research very often produces not only interesting and informative results for the social scientist, but it often obtains results that could not have been attained if overt methods had been employed. For example, Paccaguella (1997) comments on virtual communities, such as extremist political groups, who are proud of their exclusive nature and "a stranger wanting to do academic research is sometimes seen as an unwelcome arbitrary intrusion" and thus the group may not consent to the proposed research. In this instance, it may be that the only way to conduct the research is covertly. However, the moral issues relating to informed consent remain.

Benefits and harm in social research

It has been claimed that the main concern of sociology should be the search for the truth (Bulmer; 1982:10). Thus, it is argued that some level of deception is permissible in certain areas, "where the benefits of knowledge outweighs the harms" (Punch; 1984:41). However, we must assume that Punch would require that the harms should be reduced as far as possible. In this respect, the protection and confidentiality of participants is paramount. A classic example of harm caused by social research is Art Vidich's study on the Springdale community. When Vidich's study of the Springdale community was published, the identity of the community and those of individual members were blatantly clear. This caused a great deal of upset and trauma for the community, who felt betrayed by Vidich. The publication of this research in this case is generally regarded as unethical (Hammersley and Atkinson; 1995:268). It is feasible that something similar could occur in cyberspace; where researchers do not take due care to conceal the identity of the community as a whole or the identity of individual participants.

A measure to protect the community and the individual participants would be to employ the use of pseudonyms and to disguise the location of the site. This would entail, not only changing the name of the community/group, but also the names of participants, even if they themselves use pseudonyms, thus greatly decreasing the likelihood of identification (Paccagnella; 1997). On the other hand, it has been argued that this practice makes peer assessment or replication studies virtually impossible (Lee; 2000:136). Perhaps a compromise to this may be that features identifying the community/participants are eliminated from the study in publication. However,

researchers, who wish to undertake a detailed appraisal of the study or a replication study, are able to contact the author(s) for further information.

Along these lines, Jacobson (1999), in his study on impression formation in cyberspace (described in the previous section) refrained from identifying the particular MOO community. In addition to this, he used pseudonyms for participants. This generally meant providing pseudonyms for the pseudonyms that the participants themselves had chosen for his or her online activities. One negative aspect of this that Jacobson (1999:8) noted is that the online names that people chose for themselves are often very descriptive. In an attempt to keep close to the images that the names arouse, he "relied on a panel of observers to suggest names that are consonant with the imagery of those whose identities I disguise". Therefore with these measures, the community and participants are protected from harm (identification/unwanted attention) whilst the nature of the community and participation a can be openly documented.

Whilst it is extremely important to prevent harming participants in research, it is also important for researchers to prevent harm within their field – if participants / communities react against a single research project, it may in turn close the field off for future researchers. Humphreys was extremely lucky that the homosexual community embraced his study so positively. Otherwise it would have damage the field for future researchers. In contrast to this, it is certain that the Springdale community would be unlikely to accept social scientists back in their midst.

The same can be said for cyber research, if a site is too readily identifiable or perhaps if the presence of a researcher causes too great a disruption, the community may refuse access to future researchers. Lee (2000) reports a case where this occurred. A researcher, with the consent of the community, undertook a study of a MUD (multi user dimension). However, the study began to receive interest from other researchers and the community, then in the midst of a crisis period, responded by closing access (Lee; 2000:137).

This section has highlighted the many ethical problems that occur in cyber ethnography. Whilst a number of potential solutions have been put forward, on the whole, the issues remain far from resolved.

Concluding comments

This chapter has explored both the methods employed in cyber ethnography and some of the considerable ethical difficulties that emerge during such research. At the beginning of the chapter, it was noted that some commentators argue (e.g. Rice-Lively (1994) & Thomsen *et al* (1998)) that online ethnography is possible after appropriate adaptations made to the methodology. The studies reviewed suggest this is the case for all cyber ethnographic research. The principles of ethnography (e.g. extended participation to gain an in-depth understanding) remain constant whilst a few adaptations to the methods are necessary to take into account the textual nature of online communities. A number of authors, we noted, claim that cyber ethnography increases researcher reflexivity. It also brings another dimension to the researcher's interpretation, which is more readily challenged in online ethnography, owing to shifts in power away from the researcher and the open nature of such research.

The biggest challenge to cyber ethnography appears to be the absence of paralinguistic cues; this, we noted, has proven to be particularly problematic for online interviewing. Adaptations to existing 'traditional' methods have yet to be developed to compensate for their loss. The use of 'emoticons' may be one option to atone for the loss of visual cues (for example, a typographical notation ☺ to indicate smiling). Gaiser (197:141) argues that this is likely to be of little value, as "they would really need to be the sole content of a message to match their utility in face-to-face groups". This point of view overlooks the fact that 'emoticons' are commonly used in informal online communication to convey the intended meaning of a comment (e.g. to show a remark is made as a joke). Further, as people become more and more familiar and comfortable with the technology, they are likely to become more adept at online communication. So, it is possible that the lack of paralinguistic cues in online interviewing becomes less problematic.

Aside from the lack of paralinguistic cues, online interviewing holds a number of advantages and disadvantages²³. In terms of cost, time and money (travel to interview location and transcription) can be saved with this method. However, the financial cost of the equipment and the time cost in developing the necessary computer expertise (for example knowledge of HTML) may cancel out the previous cost benefits. Online interviewing is not free from sample bias, as obviously, it is limited to those that are computer literate and have access to a computer. Having said that, the scope of people that can participate is increased, as geographic

²³ Chen and Hinton (1999:table 2) provide a very useful table summarising the advantages and disadvantages of the method

constraints are not present (e.g. Steward et al (1999) cross country comparative research).

One of the most useful advantages of online ethnography is the instant transcription of conversations and interviews. This is not only time and cost effective for the researcher but also eliminates potential bias that can occur during manual transcriptions. However, Hammon (1996) reported that the printed transcripts of the online chat-room that he recorded were "very difficult to deconstruct and analyse because I could not always ascertain who was speaking to whom". It is possible this is not always a problem since it is likely to depend upon the software used (e.g. an IRC chat-room versus Firstclass Conferencing software). Nevertheless, this is an aspect that needs to be investigated before commencing the actual data collection.

One last point to note, it is likely that the advantages and disadvantages of the method changes depending upon the researcher's subject matter and working environment. For example, if a researcher chooses to study the development of a course presented electronically, (i.e. Rive-Lively's study) then it can be assumed that all participants have access to a computer and are familiar and comfortable with them. Online interviews for this group would then be off to a good start, as a number of the disadvantages (such as sample bias and non-availability of equipment), are not relevant.

It has been asserted (Singleton; 1993:349) that the use of multi-methods in data collection - observation backed up by interviews with key informants, is a good means of increasing the validity of a study. "... because of the relatively unstructured nature of field research, it is critically important to consider ways of avoiding bias and

enhancing validity". Considering the relatively new field of cyber ethnography, this seems especially pertinent. Thomsen *et al*/(1998:12) also recommends a

"multi-method approach to the study of online communities that involves the use of text and discourse analysis, a prolonged commitment to involved participant observation, and the use of qualitative interviews with group members as a means of further teasing out the 'meanings' they ascribe to the experiences of membership and participation. This would enable a multi-source, multi-method triangulation".

Another important reason for undertaking such an extended approach is noted by Kendall (1999:71), who states: "insufficient time spent on a particular forum can also lead a researcher to overestimate the anonymity of that forum". This raises issues pertaining to the validity of data as well as ethical ones.

The second section of this chapter examined some of the ethical issues that face online ethnographers. Social researchers are still at odds as to whether the informed consent of participants is necessary when studying an online community. And if consent is sought, what is the best means of obtaining this 'informed consent' – via the community webmaster/discussion group moderator or does each participant need to be contacted individually? Further, there is limited means of verifying the identities of participants (age/sex), nor can the researcher be sure that the participant has fully understood the nature of the research and thus be able to provide full informed consent. It is clear this is a matter that needs a great deal more investigation.

The identification of public and private spheres proves to be an equally problematic task in online studies. It poses additional ethical difficulties for a researcher. If a community/discussion group is considered as 'public' then informed consent may not be necessary. However, informed consent is mostly desirable within the research process – not least because of ethical imperatives. Participants have a right (moral and legal) to know they are part of a study and their behaviour/actions are being recorded. If reactivity is a consideration, then full debriefing at the end of the data collection period remains an option.

Lastly, we saw that not only does the Internet provide many new opportunities for research but it also especially facilitates covert studies. Should the researcher desire, participants need never know of his or her presence. There can be no doubt that covert research can produce very informative findings. However, it is also thought that participants and their communities, as well as future opportunities for research, are protected from harm. And this 'must' is a moral imperative.

The collection of data via both online surveys and cyber ethnography has now been addressed. Hence, we can now turn to the questions posed at the beginning of this thesis; namely: Can the established methods be transferred to the Internet? Do the tried and tested methods suffice for cyber research or are adaptations to the existing methods needed? Is the new research undertaken producing valid and reliable results? Does research, using the Internet for data capturing purposes, require major changes in the research design? It is to these questions that we now turn in the next, and final, chapter.

Chapter Four

CONCLUSION

We began this thesis by briefly outlining the history of the Internet and how the technology is being used globally by millions of people. It was also illustrated how the Internet has become an invaluable tool for academics and more specifically, for social scientists. We noted some of the obvious advantages of the Internet. These include the ability to access information in near or geographically remote databases, and to download such information for the researcher's own use. Many journals are now published online, some in addition to their print format, others entirely virtual, which also add significantly to the researcher's access to information. Another notable feature we emphasized is rooted in the ability of researchers to exchange ideas and debate specific subjects of interest with other researchers in the field. It can be used, for instance, for e-mail discussion groups on a specific topic, or for video-conferencing amongst geographically distant users. Finally, we noted that empirical data, in the form of surveys and ethnographic research, can be collected by means of the Internet. Although this new technology is still only precariously established, this has not deterred researchers from using the Internet to conduct research. They do so by using the methods and techniques of (pre-Internet) conventional research.

In Chapter Two, we examined examples of research where (quantitative) data was collected by means of Internet surveys. The adherence to correct sampling procedures is one of the most important features in the generation of valid, reliable and generalisable results. A range of conventional sampling procedures; probability and non-probability sampling, were briefly discussed. This was followed by an

investigation into how studies, which have employed the Internet for data-collection purposes, have fared in the compilation of samples.

It is worth stressing that the objective of any sampling procedure is to produce a sample that will reproduce/represent the characteristics of the population under investigation. Of course, the ideal of sampling in any form of survey research is seldom completely realised, and a certain degree of bias/error inevitably creeps in. However, if the sample is properly derived, such errors can be minimized and need not be regarded as overly problematic.

Two types of error are regularly encountered in conventional surveys. There are those arising from bias in selection. I argued that this difficulty is exacerbated when undertaking e-mail and web-based research. Firstly we saw that when undertaking an e-mail survey, there is a major obstacle of compiling the sampling frame. The data/information to be collected on a given subject is predetermined and fixed by access to and the abilities to use an e-mail address. Access is not widespread; certainly it is not as common as ownership of a telephone or postal address, which are frequently used in conventional surveys. Furthermore, we noted that to date there is not a comprehensive directory of e-mail addresses, and those lists that are available may omit certain crucial information and often will be out-of-date. Web-based surveys face even more problems, as sampling is generally non-probability and voluntaristic in nature. However, it was seen that a number of researchers, using both e-mail and web-based surveys, have employed innovative means in their attempts to address these problems.

I claimed that it is likely that these problems are due to the infancy of the technology and, no doubt, they will be overcome. Nonetheless, for the time being, it must be accepted that any proposed research using the Internet, will involve data collection from a closed sub-group, whose numbers in the population are small. And this implies that sound generalisations to the overall general population are likely to be limited.

A second problem discussed in Chapter Two, closely related to sampling issues, is 'non response'; viz. the failure for all members of the designated sample to complete (respond) to the survey. It was pointed out that there is virtually always some discrepancy between the proposed sample and the actual response rates. In conventional surveys, this can be serious enough to destroy the reliability and validity of the results obtained. In any survey, thus, it is important to minimise non-response as a source of bias. However, we saw that research has established that the proportion of the sample lost through non-response does not differ significantly between postal and e-mail surveys and that the response rates of questionnaires distributed and returned by post are notoriously low. So, in both cases, it needs to be remembered that it is very likely that there will be an insufficient basis for generalisations. Techniques have been developed in an attempt to prevent unacceptably low response rates. It was shown that follow up e-mails or letters, the option to complete the survey on paper and return by post, as well as the use of financial incentives, are generally well rewarded strategies. It was also noted that if the Internet researcher encounters an abnormally high proportion of non-response, this might be due to the fact that the questions asked are of no interest to the respondent, the questionnaire may be too long or too complex, or that unsolicited questionnaires are regarded as against 'netiquette'.

The monitoring and identification of non-respondents for web-based surveys, is, at this point, extremely difficult, if not almost impossible. Thus the section on sampling and non-response was concluded with the comment that both these issues remain open areas of bias and error in Internet surveys, particularly for web-based studies. As already stated, various researchers have been innovative in their approaches to overcome these difficulties. However, many gaps still need to be filled before sampling techniques online become as rigorous as those of conventional research.

The second section in Chapter Two dealt with issues pertaining to questionnaire design in e-mail and WWW surveys. It emerged that, in this instance, web-based surveys have an advantage over the e-mail alternative. Questionnaires placed on the WWW are less likely to be distorted by the use of different web browsers (e.g. Netscape or Microsoft Explorer). The respondents (unlike e-mail questionnaires) cannot alter the questionnaire and, they can be colourful and even interactive, both of which are useful in keeping respondent interest and so enhancing completion rates. The most obvious disadvantage of web-based questionnaires is the technical complications and researcher time associated with writing a web-form and placing it on the WWW. Another disadvantage we noted is that of the single respondent returning indefinitely many responses.

I argued that a positive option for Internet surveys would be to combine the use of both e-mail and WWW technologies. E-mail currently possesses greater control and potential in the sampling procedure and in reducing non-response. On the other hand, as we have just noted, the web is a superior medium for presentation of the questionnaire. Thus, a good alternative would be to send out e-mails, inviting

potential respondents to participate in the survey. If they wish to do so, they can click on a hyperlink (URL) that takes them to the questionnaire based on the web. Obviously, respondents would have to have access to the WWW (although in such cases, provision could be made). However to date, for whatever reason, few researchers seem to have used this method. We may conclude from all of the above that the use of the Internet in surveys is here to stay and that researchers are slowly but surely discovering its full potential.

In Chapter Three we considered how ethnographic research is undertaken using the Internet to gather data. We saw how both participant and non-participant observation is undertaken as well as online interviewing. We saw that cyber participant observation is used to obtain insights into various aspects of the Internet, such as social order virtual communities or the experiences of a particular group in the real world, e.g. young girls' perception of their bodies. Online interviewing is useful, we noted, not just as a tool to supplement the online observations but also as a tool for cross-country comparative research. It was also seen that, as is the case with Internet survey research, the sample of people that can be drawn is limited to those with Internet access. Cyber ethnography, in particular online interviewing, requires at least a basic degree of computer literacy and typing proficiency from participants. Furthermore, the lack of paralinguistic cues proved to be particularly challenging for both the participants and online interviewers. The role of the researcher, and the related issues of reflexivity and interpretation of events/situations, was then addressed. Here it emerged that many cyber-ethnographers argue that a positive aspect of cyber ethnography is that reflexivity is unavoidably heightened with this type of research.

The other main section in Chapter Three addressed some of the many ethical issues that arise when the Internet is used in social research. The first problem we considered is that of 'informed consent'. Informed consent, we noted, authorises the interviewer to act in certain ways with respect to the information she is given. Where the consent of the participant is absent or seriously compromised then the researcher's activity is thought to be morally unjustified. We saw that the gaining of informed consent in online research is problematic. As the researcher is not physically present with the participant, he or she is less able to be certain that the participant fully understands the nature of the research, nor can the researcher always verify the age, sex or geographical location of the participant.

A second moral problem we looked at concerns the unravelling of public and private spheres on the Internet. This, we noted, is an issue problematic enough in conventional ethnography. It is proving to be equally, if not more unyielding on the Internet. I argued that in contexts where this distinction is unclear; researchers are best advised to seek permission for their study before investigating or publishing his or her results. Lastly in this chapter, we considered the circumstances where it is necessary for the researcher to disguise the true nature of his or her intentions. Concealment online is particularly easy; participants can remain completely oblivious to the researcher's presence. A researcher may be concerned that revealing his or her true identity may disrupt the community (reactivity) and thus prevent the researcher from obtaining valid results. In this case, ideally, participants should be debriefed as soon as possible after observation and permission should be sought before publication of the study. However, in the event of covert research, I argued that individual

participants and the community as a whole should be adequately protected. The researcher should take care to conceal their identities. This may even mean providing pseudonyms for the pseudonyms used by participants in the community.

At the beginning of this thesis, a number of methodological questions, concerning the Internet as a data collection tool, were asked. Can the established methods of social researcher be transferred to the Internet? Do the tried and tested methods suffice for cyber research or are adaptations to these methods needed? Is the new research undertaken producing valid and reliable results? Does research, using the Internet for data capturing purposes, require major changes in the research design? It is now possible to respond to these questions. We have seen that researchers undertaking Internet based data-collection have been successful in transferring established methods to the new technology. However, in survey and ethnographical research, these 'traditional' methods do not always suffice and a number of adaptations are necessary in order to carry out research that produces results in which we can have confidence.

In response to the question of whether data collection on the Internet affects the more general issue of research design, despite the adaptations that are required to the existing methods, I can see no compelling reason for thinking that these changes require us to alter the underlying logic present in the standard research designs of survey and ethnographic research. It is possible, of course, that as the Internet develops and evolves, so will the research undertaken on the medium, and from this development, new methods may emerge and perhaps even a new research design. However this is not something that is needed at the present stage of development.

Let me conclude by drawing together a few of the threads in this discussion. As we have seen, the Internet provides social scientists with a number of new, powerful tools. Once viable solutions to sampling issues have been established, it is reasonable to assume that WWW surveys will overshadow their e-mail counterparts, owing to the formers' clear potential in questionnaire design and the ability for data to be captured directly into a database, expediting the analysis process. It also seems that online interviews hold much untapped potential. Despite the lack of paralinguistic cues, which many researchers found problematic, we have seen that the method is advantageous in many other ways. Moreover, as the WWW technology develops, it is feasible that the 'video conferencing' technology will become accessible for the average WWW users. In this event, the concern over paralinguistic cues might well become superfluous.

One further important point needs to be made. In this thesis, we have looked at survey and ethnography research as separate approaches; survey research providing an abundance of numerical data, or ethnographical research that gathers rich information from individuals via observation or interviews. These approaches, of course, are not mutually exclusive. Often the researcher, for instance, collects information from individuals within a population group by interviewing them prior to quantitative investigation by the survey method or vice versa. Triangulation on the Internet would enhance the resources of the researcher even more. For example, a research project such as Survey2000²⁴, in addition to their questionnaire, could have undertaken a number of online interviews with respondents by incorporating into the

²⁴ Witte, *et al* (2000)

questionnaire a request for respondents to indicate whether they would be willing to participate in an interview. Thus, the authors could have obtained a more in-depth understanding of their respondents to complement their numerical data. Due to the limited numbers and generally homogenous group that make up the majority of Internet users, it is obvious, however, that the Internet as the sole means of gathering data has a limited value. Therefore utilising both on and offline data collection methods appears to be another positive option. It can be assumed that through this mix of data capturing, a larger and more diverse group of participants can be reached thus enhancing the validity and reliability of the results.

There can be no doubt that the use of the Internet for data collection in social research is here to stay. Further, it is likely that the real potential of this tool has yet to be fully discovered. Thus I would recommend that further research be undertaken to explore the possibilities presented by the new technology.

REFERENCES

AAAS (1999) 'Ethical An Legal Aspects in Human Subject Research in Cyberspace: a report of a workshop' (June 10/11) prepared by the Scientific Freedom, Responsibility and Law Project, available online: www.aaas.org/spp/dspp/sfrrl/projects/intres/main.htm

Applebee, A. C. (1997) *Australian academic use of the Internet*, Internet Research: Electronic Networking Applications and Policy, Vol. 7 Issue 2

Arber, S. (1993) 'Designing samples' in Gilbert, N. (ed.) Researching Social Life, London: Sage Publications

Babbie, E. (1990) Survey Research Methods, Belmont, CA: Wadsworth

Babbie, E. (1998) The Practice of Social Research, 8th Edition, Belmont CA: Wadsworth

Baron, J. & Siepmann, M. (2000) 'Techniques for Creating and Using Web Questionnaires in Research and Teaching' in Birnbaum, M. H. (ed.) Psychological Experiments on the Internet, San Diego: Academic Press

Bauman, S; Airey, J. & Atak, H. (1998) 'Effective Use of Web Based Technology: using the Internet for data collection and communication applications', paper presented at 53rd conference of the American Association for Public Opinion Research, St. Louis, USA

Becker, H. S. (1979) 'Problems of Inference and Proof in Participation and Observation' in Bynner J & Stribley K M (eds.) Social Research principles and Procedures, New York: Longman

Bradley, N. (1999) 'Sampling for Internet Surveys. An examination of respondent selection for research', Journal of the Market Research Society, Vol. 41, no. 4 (387-395)

Bruce, H. (1999) *Perceptions of the Internet: what people think when they search the Internet for information*, Internet Research: Electronic Networking Applications and Policy, Vol. 9 issue 3 (available Emerald Virtual Library)

Buchanan, W. (1997) Mastering the Internet, Macmillan Master Series, London: Macmillan

Bulmer, M. (1982) 'Ethical problems in Social Research: the case of covert participant observation', in Bulmer, M. (ed.) Social Research Ethics, London: MacMillan Press Ltd

Burgess, R. G. (1984) In the Field: an introduction to field research, Contemporary Social Research 8, London: George Allen & Unwin

Calhoun, C. (1998) 'Community without Propinquity Revisited: communications technology and the transformation of the urban public sphere', Sociological Inquiry, vol. 68, no. 3 (373-397)

Carver, S; Kingston, R. & Turton, I. (1998) A Review of graphical environments on the World Wide Web as a means of widening public participation in social science research, AGOCG Report, available online: <http://www.ccg.leeds.ac.uk/agocg/report.htm>

Cavanagh, A. (1999) 'Behaviour in Public? Ethics in Online Ethnography', Cybersociology, issue 6, Research Methodology Online, available online: www.socio.demon.co.uk/magazine/6/cavanagh.html

CERN, (1998) 'A CERN invention you are familiar with: the World Wide Web', Available online: <http://www.cern.ch/Public/ACHIEVEMENTS/web.html>

Chen, P. & Hinton, S. M. (1999) 'Realtime Interviewing Using the World Wide Web', Sociological Research Online, vol. 4, no. 3, available online: <http://www.socresonline.org.uk/socresonline/4/3/chen.html>

Clark, R. & Maynard, M. (1998) 'Using Online Technology for Secondary Analysis of Survey Research Data – 'Act Globally, Think Locally'', Social Science Computer Review, vol. 16 no. 1 Spring 1998 (58-71)

Cohen, J. (1996) 'Computer mediated communication and publication productivity among faculty', Internet Research: Electronic Networking Applications and Policy, Vol. 6 Issue 2/3

Comley, P. (1996) 'The Use of the Internet as a Data Collection Method', available online: www.sga.co.uk/esomor.html

Coomber, R. (1997) 'Using the Internet for Survey Research', Sociological Research Online, vol. 2, no. 2 available online: www.socresonline.org.uk/socresonline/2/2/2/.html

De Vaus, D. A. (1996) Surveys in Social Research (4th edition) St Leonards, Australia: Allen & Unwin

Dillman, D. A. (1991) 'The Design and Administration of Mail Surveys', Annual Review Sociology, Vol. 17 (225-249)

Dukstra, W. & van der Zouwen, J. (eds) (1982) Response Behaviours In the Survey Interview, London: Academic Press

Elmer-DeWitt, P. (1996) 'The Information Highway: an overview', in Cozic, C. (ed.) The Information Highway, Current Controversies Series, San Diego: Greenhaven Press

Fielding, N. (1993) 'Ethnography' in Gilbert, N. (ed.) Researching Social Life, London: Sage

Fowler, F. J. (1984) Survey Research Methods, Beverly Hills: Sage

Fox, N. & Roberts, C. (1999) 'GPs in Cyberspace: the sociology of a 'virtual community', The Sociological Review, vol. 47 no. 4 November 1999

Gaiser, T. J. (1997) 'Conducting On-line focus Groups: a methodological discussion', Social Science Computer Review, vol. 15 no.2, summer 1997 (135-144)

Hamilton, J. C. (1999) 'The Ethics of Conducting Social-Science Research on the Internet', The Chronicle of Higher Education, December 3, 1999 <http://chronicle.com>

Hammersley, M. & Atkinson, P. (1995) Ethnography: Principles in Practice, 2nd Edition, London: Routledge

Hammersley, M. (1989) The Dilemma of Qualitative Method: Hebert Blumer and the Chicago Tradition, London Routledge

Hammon, R. (1996) 'The Role of Fantasy in the Construction of the On-line Other: a selection of interviews and participant observations from cyberspace', MA Sociology Scheme, University of Essex. Available online: <http://www.socio.demon.co.uk/fantasy.html>

Haralambos, M. & Holborn, M. (1995) Sociology: themes and perspectives, 4th edition, London: Collins Educational

Henry, G. T. (1998) 'Practical Sampling' in Brickman, L. & Rog, D. J. (eds.) Handbook of Applied Social Research Methods, Thousand Oaks: Sage Publications

Hine, D. (1998) 'Virtual Ethnography', IRISS '98, conference paper, available online: www.sosig.ac.uk/iriss/papers

Homan, R. (1991) The Ethics of Social Research, Aspects of Modern Sociology, New York: Longman Group

Howe, W. (1999) 'A Brief History of the Internet', Available online: <http://www0.delphi.com/navet/faq/history.html>

Humphreys, L. (1970) Tearoom Trade: impersonal sex in public places, Chicago: Aldine

Jacobson, D. (1999) 'Impression Formation in Cyberspace: Online Expectations and Offline Experiences in Text-based Virtual Communities', Journal of Computer-Mediated Communication, vol. 5, issue 1, available online: <http://www.ascusc.org/jcmc/vol5/issue1/jacobson.html>

Jones, R. A. (1994) 'The Ethics of Research in Cyberspace', Internet Research: Electronic Networking Applications and Policy, Vol. 04, Issue 3

Karaim, R. (1996) 'The Information Highway may not benefit society', in Cozic, C. (ed.) The Information Highway, Current Controversies Series, San Diego: Greenhaven Press

Kaye, B. A. & Johnson, T. J. (1999) 'Research Methodology: taming the Cyber Frontier - techniques for improving online surveys', Social Science Computer Review, Vol. 17, no. 3 (323-337)

Kehoe, C. M. & Pithow, J. E. (1996) 'Surveying the Territory: Gvu's five World Wide Web user surveys', World Wide Web Journal, vol. 1, issue 3 available online: www.w3.org/pub/www/journal/3/53/kehoe.html

Kendall, L. (1999) 'Recontextualizing "Cyberspace" Methodological Considerations for On-line Research' in Jones, S. (ed.) Doing Internet Research, Thousand Oaks: Sage

Kling, R. (1997) 'The Culture of Cyberspace: the Internet for Sociologists', Contemporary Sociology, vol. 26, no. 4 (434-444)

Kupfer, A. (1996) 'The Impact of the Information Highway: an overview' in Cozic, C. (ed.) The Information Highway, Current Controversies Series, San Diego: Greenhaven Press

Lavrakas, P. J. (1993) Telephone Survey Methods: sampling, selection, and supervision, (2nd edition) Applied Social Research Methods Series Volume 7, Newbury Park: Sage

Leaning, M. (1998) 'Cyberorg Selves: examining identity and meaning in a chat-room', M.Sc Dissertation, South Bank University. Available online: <http://www.geocities.com/Athens/Atrium/2136/Title.html>

Lee, R.M. (2000) Unobtrusive Methods in Social Research, Milton Keynes: Open University Press

Leiner, B. M. (1993) 'Globalisation of the Internet', in Lynch, D. C & Rose, M. T. (eds.) Internet System Handbook, USA: Addison- Wesley Publishing Company, Inc

Leiner, B. M. *et al* (2000) A Brief History of the Internet, version 3.31, 4th August, available online: <http://www.isoc.org/Internet/history/brief.html#Commercialization>

Lynch, D. C. (1993) 'Historical Evolution' in Lynch, D. C & Rose, M. T. (eds.) Internet System Handbook, USA: Addison- Wesley Publishing Company, Inc

Maginone, T. W. (1995) Mail Surveys: Improving the Quality, Applied Social Research Methods Series Volume 40, Thousand Oaks: Sage

Miller, H. & Mather, R. (1998) 'The Presentation of Self in World Wide Web Homepages', IRISS '98 conference paper, available online: www.sosig.ac.uk/iriss/papers/paper21/htm

Mouton, J. (1996) Understanding Social Research, Pretoria: J. L. van Schaik

Nash, B. J. R. (1995) 'Sociology Resources on the Internet: an introductory overview', Electronic Journal of Sociology, available online: www.sociology.org

Neuman, W. L. (1997) Social research methods: qualitative and quantitative approaches, Boston, Mass: Allyn & Bacon

Newsweek, (October 11, 1999) 'e-life: How the Internet is changing the World'

Nua (4th January 2000) *Internet Surveys*, Nua Limited, vol. 5 no. 1 Available online:
<http://www.nua.ie/surveys>

O' Connell Davidson, J. & Layder, D. (1994) Methods, Sex and Madness, London: Routledge

O'Connor, H. & Madge, C. (2001) 'Cyber-Mothers: Online Synchronous Interviewing using Conferencing Software', Sociological Research Online, vol. 5, no. 4, available online: <http://www.socresonline.org.uk/5/4/o'connor.html>

O'Malley, C. (1996) 'A Critical Look at the Internet' in Cozic, C. (ed.) The Information Highway, Current Controversies Series, San Diego: Greenhaven Press

Paccagnella, L. (1997) 'Getting the Seats of your Pants Dirty: strategies for ethnographic research on virtual communities', Journal of Computer Mediated Communication, vol. 3, issue 1, available online: <http://jcmc.huji.ac.il/vol3/issue1/paccagnella.html>

Paolillo, J. (1999) 'The Virtual Speech Community: Social Network and language Variation on the IRC', Journal of Computer-Mediated Communication, vol. 4, issue 4, available online: <http://www.ascusc.org/jcmc/vol4/issue4/paolillo.html>

Peters, S. (1998) 'Finding Information on the World Wide Web', Social Research Update, Issue 20, Spring, Dept of Sociology, University of Surrey

Plaut, S. M. (1997) 'Online Ethics: Social contracts in the Virtual Community', Journal of Sex Education and Therapy, vol. 22. No. 1

Pleace, N. et al (2000) 'On-Line with the Friends of Bill W: Social Support and the Net', Sociological Research Online, vol. 5, no. 2, available online: <http://www.socresonline.org.uk/5/2/pleace.html>

Punch, M. (1986) The Politics and Ethics of Fieldwork, Qualitative Research Methods Series 3, California: Sage Publications

Rea, L. M. & Parker, R. A. (1992) Designing and Conducting Survey Research: a comprehensive guide, San Francisco: Jossey-Bas Publishers

Ridener, L. R. (1999) 'The Good, the Bad, and the Ugly in Cyberspace', Social Science Computer Review, vol. 17, no. 4, winter, (445-450)

Schaefer, D. R & Dillman, D. A. (1998) 'Development of a Standard e-mail Methodology', Public Opinion Quarterly, vol. 62 (378-397)

Selwyn, N. & Robson, K. (1998) 'Using e-mail as a research tool', Social Research Update, Issue 21, Summer, available online: <http://www.soc.surrey.ac.uk/sru/SRU21.html>

Sharf, B (1999) 'Beyond Netiquette: The Ethics of Doing Naturalistic Discourse Research on the Internet', in Jones, S. (ed.) Doing Internet Research, Thousand Oaks: Sage

Sheehan, K. (2001) 'E-mail Survey Response Rates: A Review', Journal of Computer-Mediated Communication, vol. 6, issue 2, available online: <http://www.ascusc.org/jcmc/vol6/issue2/sheehan.html>

Sheehan, K. B. & Hoy, M. G. (1999) 'Using E-mail To Survey Internet users in the United States: Methodology and Assessment', Journal of Computer Mediated Communication, Vol. 4, no. 3 (March) available online: <http://www.ascuse.org/jcmc/vol4/issue3/sheehan.html>

Shipman, M. (1988) The Limitations of Social Research (3rd edition) Harlow, Essex: Longman,

Singleton, R. A. *et al*, (1993) Approaches to Social Research, 2nd Edition, New York: Oxford University Press

Smith, C. B. (1997) 'Casting the Net: surveying a Internet Population', Journal of Computer Mediated Communication, vol. 3, no. 1

Smyres, K. M. (1999) 'Virtual Corporeality: adolescent girls and their bodies in cyberspace', Cybersociology, Issue Six: Research Methodology Online. Available online: <http://www.socio.demon.co.uk/magazine/6/smyers.html>

Spink, A; Bateman, J, & Jansen, B. J. (1999) 'Searching the Web: a survey of EXCITE users', Internet Research: Electronic Networking Applications and Policy, Vol. 9 Issue 2

Steedman, P. H. (1991) *On the Relations between Seeing, Interpreting and Knowing*, in Steier, F. (ed.) Research and Reflexivity, Inquiries in Social Construction Series, London: Sage

Stern, P. (1979) Evaluating Social Science Research, Oxford

Stewart, F; Eckermann, E. Zhou, K. (1998) 'Using the Internet in Qualitative Public Health Research: a comparison of Chinese and Australian young women's perceptions of tobacco use', Internet Journal of Health Promotion, available online: www.monash.edu.au/health/ijhp/1998/12

Sunday Times (9th January 2000) 'Rabbis ban the Internet'

Swoboda, W. J. *et al* (1997) 'Internet Surveys by Direct Mailing: an innovative way of collecting data', Social Science Computer Review, vol. 15. no. 3 Fall 1997 (242-255)

Thomsen, S. R., Straubhaar, J. D. & Bolyard, D. M. (1998) 'Ethnomethodology and the Study of Online Communities: Exploring the Cyber Streets', IRRIS '98 Conference Paper, available online: <http://www.sosig.ac.uk/iriss/papers/paper32.htm>

Tse, A. C. B et al (1995) 'Comparing two methods of sending out questionnaires: E-mail versus mail', Journal of the Market Research Society, Vol. 37, no. 4 (441-446)

Wallis, R. (1977) 'The Moral Career of a Research Project' in Bell, C. & Newby, H. (eds.) Doing Sociological Research, London: George Allen & Unwin

Ward, K. J. (1999) 'The Cyber-Ethnographic (Re)Construction of Two Feminist Online Communities', Sociological Research Online, vol. 4, no. 1, available online: <http://www.socresonline.org.uk/socresonline/4/1/ward.html>

Werble, R. & Wallace, J. (1998) 'Cyber Research: the impact of the Internet on data collection', Marketing Research, vol. 10 issue 3, Fall 1998

Whyte, W. (1994) Participant Observer: an autobiography, Ithaca, New York: ILR Press

Winder, D. (1994) All you need to know about the Internet, Bath: Future Publishing

Witmer, D. F; Colman, R. W. & Katzman, S. L. (1999) 'From Paper-and-Pencil to Screen-and-Keyboard: toward a methodology for survey research on the Internet', in Johnes, S (ed.) Doing Internet Research: critical issues and methods for examining the Net, Thousand Oaks, CA: Sage Publications

Witte, J. C; Amoroso, L. M. & Howard, P. E. N. (2000) 'Method and Representation in Internet-Based Survey Tool- Mobility, Community, and Cultural Identity in Survey2000', Social Science Computer Review, Vol. 18 no. 2 (179 - 195)

Yun, G. W. & Trumbo, C. W. (2000) 'Comparative Response to a Survey Executed by Post, E-mail, and Web Form', Journal of Computer Mediated Communication, Vol. 6, no. 1 (September) available online: <http://www.ascuse.org/jcmc/vol6/issue1/yun.html>