International Journal of Internet Research Ethics 1.1

Editors-in-Chief

Elizabeth A. Buchanan, Ph.D.
Center for Information Policy Research
School of Information Studies
University of Wisconsin-Milwaukee
buchanan@sois.uwm.edu

Charles M. Ess, Ph.D.
Distinguished Research Professor
Drury University
cmess@drury.edu

Annette N. Markham, Ph.D.
Senior Research Fellow
Center for Information Policy Research
School of Information Studies
University of Wisconsin-Milwaukee
amarkham@gmail.com

International Journal of Internet Research Ethics 1.1

Table of Contents:

INTRODUCTION: Why IJIRE?

The *International Journal of Internet Research Ethics* represents a qualitatively new phase in the increasingly extensive and increasingly urgent deliberation upon and development of ethical guidelines for Internet research.

This deliberation is perhaps as old as Internet research itself – and has grown rapidly alongside the Internet. Less than ten years after the first set of four World Wide Web servers came online in 1992, the Ethics Working Group was established by the Association of Internet Researchers in 2000, charged with developing ethical guidelines for Internet research. There were multiple factors driving this decision and work: especially within the United States – where, for better *and* for worse, much of the Internet research of the 1990s originated – researchers in a range of disciplines were encountering a consistent series of difficulties as they submitted their research proposals to their university Institutional Review Boards (IRBs). Such IRBs are mandated by U.S. law to ensure that human subjects research undertaken with federal funds does not violate basic requirements of human subjects protections, as codified in the Code of Federal Regulations, and as described in the seminal *Belmont Report*. Informed consent, respect for persons, and beneficence are fundamental, grounding elements of human subjects research.

While there is no quarrel but that such protections are necessary, just how these protections were to be applied to research on "virtual subjects" in online environments was not always clear, neither to IRBs nor researchers. And even without the promptings of an IRB (or its institutional equivalents in Australia, Canada, and the U.K.), thoughtful and reflective researchers across the globe found themselves confronted with difficult ethical questions evoked in their research for which no obvious or immediate guidance was available. There was hence an increasingly widespread recognition that ethical guidelines for Internet research were urgently needed.

The AoIR Ethics Working Group published their ethical guidelines in 2002, following the approval of the AoIR membership – and, happily, those guidelines have found extensive use both within the United States and around the world. But since that time, of course, both the Internet (now connecting over 1/6th of the world's population)

and thereby the scope of Internet research, has continued to dramatically expand. At the same time, as we learned from the most recent ethics panel at the AoIR Internet Research 8.0 conference in Vancouver, IRBs and their correlative institutions in the U.K. and Australia are no longer an Anglophone phenomenon. Rather, several of our European colleagues reported to us that they are increasingly facing the kinds of questions and review procedures that until recently, seemed to be a primarily US phenomenon. This means that, just as the Internet continues to expand its global reach, and as Internet research follows close behind, becoming an ever-more international and interdisciplinary enterprise, international perspectives on the myriad ethical issues that inevitably emerge when we seek to study human beings will continually grow urgent.

In our view, the scope and depth of reflection and scholarship on Internet research ethics has reached the point where a journal devoted exclusively to these topics is justified – and, indeed, necessary. It is not simply that more and more people, in more and more places around the world, are researching engagements online by more and more people, who likewise reflect an ever-increasing diversity of countries and cultures; it is that *IJIRE* is needed as a central venue for the collection and dissemination of a rapidly burgeoning body of research and reflection. It is further the case that as our ethical difficulties and (re)new(ed) insights into their possible resolutions exponentially increase with the global growth of the Internet, researchers, the people they study, and the institutions responsible for Internet research oversight thus ever more urgently need the reflections, understandings of law and professional ethics, and insights into new ways of resolving sometimes new ethical difficulties that we seek to make available through *IJIRE*. This journal is meant to bridge theory and practice. We seek to represent disparate disciplines, locales, ideologies, and legal perspectives. We have assembled an amazingly unique editorial board, top scholars in the field of Internet research ethics, representing an array of nationalities and disciplinary expertise. We embrace writings from students, and encourage feedback from academic communities across the globe. We recognize cultural and disciplinary differences in writing, and will not "force" our authors to fit any rigid model or mode of presentation. In this first issue, we present an array of articles, which explore various dimensions of Internet research ethics, ranging from methodological to legal to philosophical. Each article demonstrates the ongoing complexities researchers

and policy makers are facing—and indeed, embracing—as Internet research and the consideration of its ethical implications assumes greater prominence across the world.

In addition to our editorial comments and regular articles, each issue will feature "small talk," a short conversation with a leading voice in Internet research ethics. For this inaugural issue, we feature Annette Markham, who recently held a conversation with a graduate seminar, Global Issues in Internet Research Ethics, at the University of Wisconsin-Milwaukee. Next issue will feature Susannah Stern. A forthcoming feature will include a practical scenario, or ethical dilemma, that a researcher or board has faced in their experiences in Internet research. Feel free to send in your experiences (ijire@sois.uwm.edu). Again, we hope *IJIRE* reaches a range of stakeholders, through a combination of theory and practice. We are bound to have a few mistakes along the way, and we ask in advance for your understanding!

Welcome to our inaugural issue!

Small Talk: Annette Markham

Annette Markham is well known for her important publication, Life Online: Researching Real Experience in Virtual Space (Alta Mira, 1998). Her forthcoming work is Internet Inquiry: Conversations about Method (co-edited with Nancy Baym).

We got some insight about interviewing and identity issues from Annette in April 2007, when she guest lectured in a graduate course at UW-Milwaukee on Global Issues in Internet Research Ethics. This is an informal excerpt from our discussion.

QUESTION: You mentioned in your book that online interviewing can be frustrating because there is really no way to see the body language, mannerisms, or pick up on verbal cues from the person being interviewed. Do you think this is a significant drawback and/or detrimental to the information gathered in online interviewing as opposed to face to face interviewing or phone interviews?

This is an excellent question because it goes to the heart of interviewing in general, not just interviewing online. Although it might seem to be a straightforward question, it raises HUGE issues, methodologically and ethically speaking. First, the short answer: No. In this case (my book), there was no drawback or detriment.

But, let me specify that response a bit: I wasn't looking for information that would be embedded in the body. In fact, that would have detracted from the type of information I really wanted. I wanted to know them in their online identity. If they had to come offline, this would be a different sort of identity, yes? I'm not saying they're necessarily separate, as if there is some offline versus online identity. But in any context, we adopt or display different masks, consciously and unconsciously. Researchers need to consider this carefully. If I want to interview people about their night-time college life, I won't interview them when they're visiting their family for Christmas. I'd rather interview them at night, at college.

Having said that, let me expand on the issue somewhat: For me, there are some other important questions embedded in your question above:

1) We don't see embodied characteristics...so what DO we see in an online interview?

2) What do embodied characteristics provide us in the interview process?

3) What is privileged (or conversely, marginalized) when we rely on embodied ways of "knowing" others (as researchers)?

Most of the time, in face to face interviews, we researchers take a lot for granted; body posture, movements, eye contact, etc. We notice these things, but not always at the level of consciousness. Yet all this embodied 'stuff' influences the researcher, whether or not he or she pays attention to how or why.

If we notice posture, movement, mannerisms, etc., we'll use this information to help us interpret or categorize the participant. But what about those things we don't notice? What about those things we don't notice at the surface level but use at a subconscious level? In a physical interview, for example, we might sense a subtle body odor that makes us want to spend less (or more) time with the interview subject. Or someone's face might unconsciously remind us of a bully we knew in the 2nd grade, which might impact how we ask questions.

What if we use these nonverbal/paralinguistic signals as part of the interpretation but we're completely off-base? We are taught that the eyes are the window to the soul, that body language doesn't lie, that you can tell how someone is feeling by looking at them. This privileges our five senses as the most accurate way of knowing someone else. For the most part, we believe that we are capable of understanding what those nonverbal signals mean. I've been proven wrong so many times I now believe THIS might be the larger detriment to interviewing.

Basically, qualitative methods textbooks don't talk much about the nonverbal aspects of interviewing. Perhaps it's because it is assumed that we know how to handle this stuff. Wow, what a mistake!

There's even less discussion of paralinguistic aspects of online interviewing. Online interviews remove a lot of markers (socioeconomic markers) like race, gender, class. We don't see the clothing style, famous label sunglasses, any obvious racial characteristics, and gender. But we see other things that will "mark" the participant, like typing speed, grammar, precision, sense of humor (or lack thereof), type of personal web

space, deliberate verbal characteristics, etc. So the cues we use to help us make sense of who the person is? These markers still exist, but take different forms.

Is it important to have the embodied information? Yes and no. Yes, because it is a comfortable and familiar way for us to believe we "know" the other person. But in terms of qualitative research, what do these signals convey? How accurate are these signals in everyday life? Sometimes not very accurate at all, turns out. Examples: When a person walks into an expensive boutique, the employee who dismisses this person because she or he is not dressed in the 'right' clothes may be making a big, expensive mistake. People tend to attribute criminal intention more to African Americans than White Americans, a massive mistake based only on appearance.   People tend to think homeless people are dangerous or stupid, another mistake in generalization.

So back to the issue of qualitative research. If we want to know more about social life, we use observation and interaction to learn more about people. The snap judgments we make based on nonverbals are suspect from the start because we don't have very reliable interpretive filters. We use generalizations and stereotypes and habit to interpret what these nonverbal signals mean. This can be accurate, but can just as easily be a big mistake.

 I'll just add one more thing. Each research situation is unique and the method for getting data should be determined based on this situation. So, even though I preferred/needed online-only information, there are research situations that warrant a mixture of face to face and online interviewing. A scholar at London School of Economics, Shani Orgad, discusses this issue extensively in her work re: online breast cancer support groups.

QUESTION: We found the dialogue in chapter 2 of your book interesting when you changed your name to Bambi and altered your personality. You also talked about how this type of behaviour can give users a rewarding sense of control. Do you think this is a healthy perspective or could this have negative consequences related to real life social interactions?

Experimentation with other/alternate identities can be helpful.   Here's an example: I know a person who was extremely uncomfortable and shy around men. Couldn't speak,

was scared to be around men, seemed to always be taken advantage of, and so on. On a fluke, she created a male avatar in an online community. She later told me that in a really short period of time she began to lose her fear of men. Online, she interacted with other male avatars, protected by a shell of masculinity (her avatar). She could experiment with conversation and interactions. She could observe them, be with them, without being in her own (female, scared) body. She translated this back to her embodied presence, started experimenting in face to face encounters.

If people feel locked in the body or identity they have, experimenting with identity online can be a form of survival, a means of escape, or it can be a trap. In small rural towns, young people often feel like identity is predetermined by parents, teachers, and friends. There's little opportunity to be something else. There's no anonymity. For some people, this can be comforting, because they know who they are, their identities are reflected in the mirrors of other people's perceptions (this is not my idea; this comes from the looking glass theory of identity). For other people, it can be a prison because they can never find what they perceive might be their true identity (not that I believe there's any such thing as a 'true' identity...but we like to believe there is).

Negative side to all this: One's comfort about one's identity does not come solely from the self. It comes from others. If people need to be online to feel good about themselves, this is probably a signal that something is not right in the physical world. And, after all, we live in the physical world, so we have to confront, be comfortable, and love what's in front of us in the mirror every day. Otherwise, what do we have? This problem is not restricted to online/offline identity management. Having said that, I will note here that the vast majority of people I talk with understand this issue and they're not blind to the idea that they have to live in the physical world. So the fears of the mid-90s (internet addiction, loss of social skills) appear to be overstated.

Your question here raises one other point: control. Whether online or offline, when it comes to communication and the "presentation of self" to others, we think we have more control than we actually do, anyway.

People spend a lot of time engaging in what we call image management: it's obvious when you see how people wear certain clothes to portray a certain image: for

work, parties, parents. We can see this happening in the classroom; students deliberately sit in the front row of class (or the back), they slouch or sit attentively. We see deliberate image management it everyday behaviours: We engage in flirting, pious, respectful, humble, etc., etc., types of communicative behaviours. We perform this image management stuff because we think our behaviours will have particular effects. And we tend to overestimate the extent to which other people get the exact message we intend to convey through these behaviours. Considering how much we also believe we are misunderstood, our control is an illusion. Thankfully, we have this illusion, really. It enables us to get through the day. Also, since everyone has this illusion, we end up with a shared belief that we understand each other.

Incidentally, it was hilarious and fun being Bambi. I could never have been a "Bambi" in physical life. Nobody would believe it. It felt like putting on a wig and changing my entire personality. I couldn't do that with ease in the physical world, but I achieved some of the effects by simply changing my name online. How bizarre is that? All because of a name.

....Of course, I was engaging all sorts of terrible stereotypes to experiment or worse, amuse myself, stereotypes that a real person named Bambi would probably not appreciate. And I was perpetuating stereotypes by interacting with others using this exaggerated 'Bambi' identity...there's some potentially significant damage done in the larger social picture there.... But that's a discussion about ethics for another day....

Ethical Approaches to Robotic Data Gathering in Academic Research

G. N. Allen
Information Systems Department
Marriott School of Management
565 Tanner Building,
Brigham Young University
Provo, UT 84602 USA
gallen@byu.edu

D. L. Burk
University of Minnesota Law School
229 19th Ave South
Minneapolis, MN 55455 USA
burkx006@umn.edu

*C. Ess**
Philosophy and Religion Department
Drury University
900 N. Benton Ave
Springfield, MO 65802 USA
cmess@drury.edu

Abstract

Internet researchers increasingly have at their disposal of an array of automated software agents, or "bots," which can rapidly and efficiently retrieve a variety of economic and technical data from publicly accessible web sites. While these automated tools greatly facilitate the retrieval and analysis of data for academic research, they may pose ethical problems for Internet researchers. Specifically, automated software bots place some load on servers being accessed, possibly in contradiction to the expected use of such servers, and possibly in violation of the legal prerogatives of web site owners. Determining how and when to access such web sites, and whether to seek the consent of web site owners for retrieval of publicly accessible data presents an apparent conflict between general principles of information policy and the emerging legal precedent regarding trespass to computers. This conflict may be characterized as pitting utilitarian considerations against deontological considerations in a fashion reminiscent of previous debates over informed consent in on-line research. In this paper, we examine both utilitarian and deontological characterizations of the ethical obligations of researchers employing automated data retrieval bots, and argue that the contrasts between the two approaches do not necessarily

---

* Dr. Ess is an editor on the IJIRE. As a co-author of this article, he recused himself from the review or selection of the article.

result in conflict. Instead, we argue that the tension within the relevant practices indicates the need for a "meta-choice" between utilitarian and deontological considerations. We further suggest certain factors that may differentiate such a "meta-ethical" choice in the context of automated data retrieval from the "meta-ethical" choice presented in previously identified contexts of human subjects research or of web browser technology design. In the end, we argue that by analyzing the ethical issues in terms of the contrast between utilitarian and deontological ethics, it is possible to resolve some of the ethical dilemmas regarding automated data retrieval in fruitful and cogent ways.

Introduction

The growing popularity of the Internet as a means of communication has significant ramifications for academic researchers. Not only does the Internet facilitate traditional research efforts by enabling remote collaboration, it also provides new ways to accomplish old tasks and opens important new avenues for research. Moreover, the Internet, and the behavior of individuals and firms interacting with it, has itself become an area of academic research. As researchers study the behavior of individuals and organizations in electronic environments, the Internet itself facilitates access to large amounts of on-line data generated by these activities. Examples of such data sets include the prices and offerings of competitors in electronic marketplaces and the behavior of both sellers and buyers in electronic auctions. The amount of data available on the web pages of target sites leads researchers to look for automated procedures for data collection. Not only does an automated approach allow larger amounts of data to be collected more accurately, and in a shorter period of time than can be accomplished manually, it also provides for the reliable gathering of longitudinal data. Automated Internet data collection agents have recently been used by several researchers in various domains (e.g., Clay, Krishnan, & Wolff, 2001; Clemons, Hann, & Hitt, 2002; Baye & Morgan, 2001; Pan, Ratchford, & Shankar, 2002; Shankar, Ratchford, & Pan, 2002; Kauffman & Wood, 2000; Hahn 2001). Much of the research facilitated by these agents would be impossible or impractical without such tools.

However, at the same time that they open and facilitate new avenues of research, automated data collection agents present new ethical challenges, as the features of

automated agents that make them most appealing for research use also raise issues as to their impact upon targeted web sites. Although the ethics of on-line research has gained increasing attention by individuals, governments, and professional organizations (such as AoIR, Association of Internet Researchers), we are to date unaware of any literature addressing the ethical standards for use of automated agents. However, the salient features of automated data collection have been the subject of legal scrutiny, particularly in the United States. Consequently, because law frequently incorporates moral or ethical norms, in this paper we examine the relevant legal precedent for indications as to the ethics of automated data retrieval.

In doing so, we encounter an apparent conflict within the applicable current legal standards and researcher's response to those standards. From the standpoint of applied ethics, two different sorts of arguments appear here – i.e., *utilitarian* and *deontological*. In contrast to a utilitarian cost-benefit approach, deontological approaches emphasize the primary – and (more or less) absolute – importance of basic rights, duties, obligations, etc. In the following, we explain more carefully what these approaches entail in the context of automated data collection, and then analyze the conflicting approaches vis-à-vis the emerging legal positions, specification, and practices. Doing so may help us evaluate this conflict on a second, perhaps more fruitful level. By analyzing the divided approaches within this larger framework, we hope to show that the apparent conflicts between the various emerging law and practices, is in an important sense just that – *apparent*. From a larger perspective, the apparent conflict more fundamentally reflects a deep difference between general principles of legal doctrine and research policy as utilitarian, whereas the positions of emerging law, specifications, and practices may be viewed as deontological. In this light, the conflicts between these two positions require a second level "meta-ethical" debate over which of these two positions should be give more weight in the case of conflict.

We begin by describing the nature of the research tools employed in automated data retrieval and the types of research implicated by these tools. We then outline the ethical problem presented by this type of research, and review the legal precedent that might be applied to this research activity. We note in particular that legal policy and practice in these circumstances is deeply divided between prescriptions that may be

characterized as, respectively, utilitarian and deontological in nature. This leads us to examine the extent to which legal doctrine may be relied upon to guide or indicate good ethical practice in these circumstances.

Automated Data Retrieval Robots

Automated Internet data collection agents are simply computer programs designed to interact with web servers via the Internet to collect various kinds of data (Kauffman, March, & Wood, 2000). Because they retrieve information on behalf of humans and operate in an environment designed for human interaction, they are often referred to as software agents, or simply agents. Computer programs that automatically retrieve information from Internet web sites have been called robots (or "bots" for short). For several reasons, we shall refer to such programs as bots or robots through the remainder of this discussion, and avoid the term agent although it is quite common. We believe that term "agent" is significant and somewhat loaded for at least two and possibly three reasons, relating to both connotational and denotational valences of the term as it has been used in other contexts.

The first of these derives from the legal concept of agency. Under the law of agency, an individual, or principal, may designate and authorize another to act on his or her behalf, managing the principal's property or interests, entering legally binding relationships, and for that matter incurring liability as if the principal had acted personally. A large body of law has developed to define and govern the use and misuse of such delegated action. Human agents under the law of agency are typically expected to act according to the instructions of the principal, effectively as extensions of the principal's will. Of course, having their own independent will, legal agents do not always act precisely according to instruction, which may sometimes work to advantage to the principal, for example when unforeseen circumstances arise. This may also work to the principal's detriment when the agent makes a mistake or exercises poor judgment within the scope of his authority. Or, the agent may act in his own interests rather than that of the principal, and a good portion of the law of agency considers the redress available to the principal when agency is misused.

Software agents at the present state of technology typically do not behave independent of their programming, although they may of course malfunction. But we see no reason that the lack of independent behavior would necessarily preclude software robots from being legally regarded as precisely instructed and faithfully executing agents. Indeed, as electronic data interchange and other form of automated commerce have become more common, the legal consequences of programmed commercial instructions have become routinely imputed to whoever deployed that program.

In a different, non-legal sense, the term "agent" may also refer to a discrete causal entity. The epidemiologist may speak of etiological agents, or the historian of agents of change. In this sense, agents may be sentient and animate, or they may be insensate and inanimate. Volitional action is not necessarily within the valence of the word in this sense. Such a view of agency likely reaches its apex in actor-network theory (ANT) where the causal agents, or actants, contributing to a particular social phenomenon are not sorted by volition or sentience; neither human nor non-human causes are privileged in the causal analysis.

Software robots may surely be agents in this sense of the word, whether or not they are legal agents. But we also note that an entity, or causal agent, need not be an agent in the legal sense to incur liability or other legal responsibility on behalf of another. The owners of mobile but non-living agencies, such as confined waters on a landowner's property, or noxious fumes, may be held liable for the damage done by such causal agents when those agents are accidentally or purposely released. Similarly the owners of living but non-sentient agents, such as livestock, may be liable for the damage done by such agents when they are released or escape confinement.

It may well be that software robots might be said to act as agents in these two senses of the word, but neither is necessary for our analysis here. In particular, we do not propose to explore the legal agency of such robots, given that legal agency is not necessary to incur liability. We sidestep these issues, and for that matter the term "agent" in part because an analysis of legal agency is not our purpose here, but more because we are aware that lurking behind each of these uses of "agency" lies a third use of the term agent, having a connotation of free will or independent "free agency." This is a far more

complex question going to questions of cognition and personhood. We recognize that a sufficiently complex software robot might either possess agency in this meaning of the term, or behave in such a manner that was indistinguishable from the exercise of such agency. It is less clear that a software agent of the sort we are considering is sufficiently complex to rise to this level. In this paper we focus on the ethics of programming and deploying software robots as the technology currently exists, which means that we focus on the ethical situation of the programmer or researcher using the technology. We expect consideration of the software robot as an independent ethical entity will someday come, but today is not that day.[1]

Whatever their potential legal or moral status, the software robots we consider here are programmed to retrieve web pages and parse them to find data to be stored for analysis, for references to images to be downloaded, and for links to other web pages that might contain useful information. Because such data collection bots have the capability to simulate data being entered into a web form, such as a customer order form, and posted by a user, they can dynamically interact with electronic commerce web servers and collect detailed data about various practices and behaviors in the online environment. Many programming environments provide tools for the development of such agents, so they are becoming increasingly easier to develop and deploy in meaningful ways (Allen & March, 2000).

Data collection bots vary widely in their architecture and sophistication. They can retrieve documents in either a serial or parallel fashion. When an agent data collection bot requests a page from a web server, it typically takes much longer to retrieve the page than it does to parse into data elements and process it for the data of interest. If a data

---

[1] For further discussion of the complex of meanings surrounding the term "agent" – especially within the context of whether or not computers and computer programs can be said to exercise *moral* agency, see especially: Frances Grodzinsky, Keith Miller, and Marty J. Wolf, "The Ethics of Designing Artificial Agents"; Deborah Johnson and Keith Miller, "Un-Making Artificial Moral Agents"; and Alison Adam, "Ethics for Things", forthcoming in a special issue of *Ethics and Information Technology,* on "Luciano Floridi's Philosophy of Information and Information Ethics: Critical Reflections and the State of the Art." Adam, in particular – who likewise draws on Actor-Network Theory (ANT) as well as on Floridi's Philosophy of Information (PI) and Information Ethics (IE) suggests a helpful distinction between, say, dogs and seat belts as moral *agents* (that is, they act in ways that can be morally evaluated in terms of good and bad), but not necessarily as morally *responsible* agents (they are largely carrying their "programs" as either trained and/or designed by human beings to achieve specific ends).

collection bot agent processes its requests serially, that is, waiting for the first request to be retrieved and parsed before the second request is issued, it places a much lower burden on the web server with which it is interacting than an agent that issues requests for a large number of pages with no delays and parses them after they have all been received. Data collection agents' bots can also be built to operate in a distributed manner in which several copies of a bot agent can be running simultaneously on different computers, all working together to complete the same data collection task. Because these bots are designed to interact with web servers over the Internet, they are easily adapted to cooperate with each other over the same channel, making it relatively easy to deploy massively parallel, geographically dispersed distributed data collection networks with hundreds or thousands of nodes.

Data collection bots for academic research typically have a predefined set of data to collect. This set may be defined by a specified set of Internet sites or URLs as targets for requests, a particular description of a set of products, such as a set of ISBN numbers, or by some third-party list, such as "the Billboard 200" list of popular music recordings. In any case, the set of desired data items are specified in an unambiguous, targeted manner. Data collection bots are often deployed to gather data that is changing with some degree of regularity. Data collection bots have some important differences from resource discovery bots. For example, a resource discovery bot may be used to discover and index the page that holds price information for a particular product being offered by a particular vendor. In doing this, the resource discovery bot seeks the location of the page, but is not concerned with the actual price, which may change frequently. The data collection bot is likely to be used to collect the price at specified intervals so a researcher can gather a longitudinal data set for use in analyzing Internet pricing. This requires that data collection bots make frequent return-visits to particular documents.

When academic researchers deploy automated data collection bots, they can collect very large amounts of data in relatively short periods of time, but this power comes at a cost—a cost only partly borne by the researcher. Both data collection bots and resource discovery bots employed by search engines such as Yahoo, Google, or Altavista, use the resources of web servers in ways that may not have been intended by the owners of those resources. Repeated interaction with the web site being accessed places some

load on the equipment of the owners or hosting agent of the web site. In particular, deployment of automated data collection bots allows significantly more requests than would be feasible by individual consumers or Internet users requesting the data manually. As a result, there has been some discussion (e.g. Kostner, 1995) regarding the appropriate place of such data collection agents in the Internet community.

Foundation for Ethical Concern

The expectations of individuals posting information on the Internet are critical to this analysis. When individuals or organizations make information available through the World Wide Web, absent some type of password or access protection, they are making that information accessible to the public. As with other means of putting information before the public, such as publishing a book or making a television broadcast, there is some cost associated with placing the information in a particular format. However, with such traditional channels, incremental costs of accessing information are borne solely by those retrieving the information as individuals buy or share books and as they buy necessary hardware to receive a television broadcast. This is not the case in an electronic environment. After individuals and organizations invest in the necessary computer hardware or subscribe to the necessary services to be able to post information on the World Wide Web, there is a continual and incremental cost associated with responding to individual requests for information. These response costs are not simple variable costs that vary directly with each use. They are step-variable in nature, so that a single added request adds almost nothing to the total cost, but a significant increase in accesses requires a significant step up of response facilities. In other words, the infrastructure necessary to support 10,000,000 requests per day is more expensive than that required to support 1,000 requests per day. Thus, although the requester of information pays part of the incremental cost of processing a request for data by spending resources to make the request, the supplier of the information pays the remainder of the cost in answering the request.

There are a variety of reasons that those placing information in an electronic public arena are willing to bear both the up-front and incremental costs. It may be because they want to change public opinion, or that they hope to increase sales (either

directly through their electronic storefront or through their traditional channels), or it may be just because of an altruistic desire to help humanity. Whatever the reason, it is almost certainly based on the premise that when a request is served, a human reader will be exposed to the information on the page, or that there will be some other positive externality of the interaction. For example, when an Internet search engine such as Google, Altavista, Excite, or Yahoo deploys a resource discovery bot that requests a page from a web server, the individual or organization that ultimately bears the cost for responding to the request hopes to gain a second-order benefit. Although no human will be directly exposed to the information served as the response, it will be indexed and ultimately made more accessible as individuals querying the search engine become aware of the document contents and are referred back to its source. Thus, when a search engine employs a bot to gather data from various websites they consume some resources of the site, but they provide second-order benefits in return.

This is not the case when an academic researcher deploys a data collection bot. As the bot executes its task, it consumes the resources of various web servers, and typically reciprocates with virtually no benefit. Although there may be a second-order benefit from published research that brings awareness to a particular website, there is no guarantee that such publicity will be favorable to the site. Accordingly, the use of Internet data collection bots constitutes a form of social free riding—the taking of a benefit without paying the associated cost. Significantly, this cost may not have been the cost anticipated by the site, or a cost that the purveyors of the site would necessarily be willing to bear if asked.

Researchers may of course allay such concerns by requesting permission to access the site, allowing the owner or operator of a web site the opportunity to voluntarily assume the additional load that comes from academic data retrieval. However, such requests have their own costs, slowing the pace of research and possibly negating the advantages of data retrieval automation. Additionally, in much the same way that people behave differently when they know they are being observed, websites may present different information if they know they are being examined for purposes of an academic study. If there is reason to believe that a site, once aware of the study, would tailor the information sent to an academic researcher, it may contaminate the research collection to

notify the organization of the study through a request for express consent. And of course, some web site owners may well decline to grant permission for research activities that will accrue little or no benefit to them. Thus the academic researcher faces a dilemma: even though the data collected is ostensibly publicly available, the method of collection may well raise an ethical concern and the most obvious method of allaying that concern—consent—may well seriously hamper research efforts.

Established Legal Precedent

The problem of free riding on Internet servers has not gone unnoticed in commercial contexts, and courts have already begun to formulate legal standards for such activity (O'Rourke, 2000; Elkin-Koren, 2001). These decisions have sounded in a range of legal doctrines, including federal and state statutory prohibitions against "unauthorized access" to networked computers. Copyright claims have been less successful. Other, as yet untried legal theories could be imagined as the basis for website exclusion, such as theories of unfair competition, or, in the EU, statutes protecting database rights. But to date, by far the most successful legal claim asserted against on-line automated data gathering has been based on a theory of trespass to chattels, that is, to moveable property. Several courts have now embraced this renovated legal theory, revising its classical elements into a new type of tort claim adapted to the context of the Internet.

At common law, a claim of trespass to chattels required interference with or dispossession of the chattel, resulting in some harm or damage to the chattel, or pecuniary loss to the owner. This definition of trespass has been reformulated for networked computers to hold that electrical impulses satisfy the common law requirement of physical contact, and the increased load on the networked system qualifies as interference or dispossession. Harm or pecuniary loss is often presumed from the loss of processing cycles or the diversion of data storage capacity. For example, in *eBay v. Bidder's Edge*, a United States District Court relied on a theory of trespass to enjoin Bidder's Edge, an aggregator of on-line auction data, from automated collection of data from the eBay site. Similarly, in *Register.com, Inc. v. Verio,* a trespass theory was used to penalize the automated collection of ownership data from a publicly accessible domain name database. Most recently, in *American Airlines v. Farechase*, a Texas state court enjoined

the producer of software used to search in real time for air fares on airline websites, on a theory of "contributory trespass," that is, aiding and abetting others to trespass on the airlines website.

Each of these decisions has hinged in large measure upon the question of notice, and of authorization. Harmful contact with the chattel only constitutes a trespass if the contact is unauthorized, and at least some types of contact with computer servers can be inferred to have been implicitly authorized by interconnection with the Internet. Why else connect the server to the network, making the site publicly available, unless contact, at least some type of contact is desired? Like any other license, implied licenses have some limit, and can be revoked or overridden by an explicit license. The question then becomes the extent of the implied license to access the server, and the ability of the owner to explicitly limit or revoke such implied authorization.

Courts have inferred such limits, or alternatively found such revocation, in the actions of the site owner or in the terms of certain publicly available documents. Specifically, terms of service posted on the website, forbidding certain types of access, have been held to constitute evidence of limitation or revocation. Similar evidence has been found in the presence on of the "robots.txt" file on a web server. Written in conformance with the "Standard for Robot Exclusion" (Kostner, 1995), site of the "Standard for Robot Exclusion," contained in a machine readable "robot.txt" file. This file, intended to limit crawling by resource discovery robots, can indicate to a properly programmed robot which pages the site owner wishes indexed and which pages the site owner does not want indexed. Robots need not be programmed to honor such automated requests, but their presence, and an indexing industry norm of doing so, has been taken as explicit permission or denial of site access.

While the courts have almost universally embraced some form of exclusionary right for website owners, legal commentators have for the most part been highly skeptical of this trend. The initial and ongoing criticism of these cases has been largely based upon instrumental grounds: that the legal result will have deleterious effects on Internet activity, that it will result in an undesirable fragmentation of property rights on the network, and that it conflicts with policy goals of U.S. intellectual property law. Indeed,

one of us (Burk, 2000) has explicitly advocated that a legal cost-benefit analysis should be incorporated into the determination of such cases. While a minority of commentators has applauded the development of these cases, their applause has also been based upon instrumental grounds: praising the assignment of property rights in web sites as necessary to the development of healthy licensing markets and efficient allocation of resources (Epstein, 2003; McGowan, 2004).

A fairly clear example of such public policy occurs in the copyright area, where the United States Supreme Court has made clear that the U.S. constitution forbids intellectual property rights in unoriginal works, such as certain factual compilations, no matter how great the harm that may result to an individual's interests from the refusal to extend such protection. In the seminal *Feist v. Rural Telephone* case, the United States Supreme Court held that copying of a telephone book's white pages is not only permissible under copyright law, but that copyright law cannot be extended to prevent such copying even should Congress wish to do so -- despite the effort and investment that might have gone into creating the telephone book, it is free for the taking. The constitutional public interest in access to factual information is too great to allow such access to be restricted by copyright. Fair use offers another example from copyright where we allow unconsented use of another's intellectual property, even to that individual's detriment, in the overall interest of the public. This suggests that public policy dictates that publicly accessible data on web sites should be similarly appropriable. Indeed, a number of scholars are highly suspicious of the "trespass to computers" line of cases, as they appear to be an attempt to make an end run around copyright law, seeking protection for published data that could not be protected under copyright.

Law as a Moral Compass

In assessing the ethical implications of these arguments, we notice first note that they are distinctly *utilitarian*. Briefly, utilitarian ethics emphasize a kind of moral cost-benefit approach, so as to ask the question: do the probable benefits of an act outweigh its probable costs, in which case, it is morally justifiable – or do the costs outweigh the benefits, in which case, the act is *not* morally justifiable? Indeed, the trespass cases themselves contain a good deal of language weighing the costs and benefits of

recognizing an exclusionary right for website owners. This stems in part from the requirement of harm or impairment as an element of the tort, but to an even greater extent from the procedural posture of the decisions, which were largely concluded on motions for a preliminary injunction. Such motions occur very early in the litigation process, before complete information is available – for example, before there has been a complete determination of the rights of the plaintiff. Due to the incomplete information available to the court when granting preliminary relief, the calculus of factors in deciding the motion includes weighing the likely injury to the plaintiff if the injunction is improvidently denied against the likely injury to the defendant if the injunction is improvidently granted.

This characterization is important to observe for several reasons. First of all, it is characteristic not only of U.S.-based approaches to intellectual property, but of U.S.-based approaches to research ethics – in contrast with the more *deontological* approaches characteristic of Europe and Scandinavia, in both the areas of research ethics and of intellectual property (Burk, 2006). Thus, the general policy of American information law, together with the practicalities of research practice, might lead one to conclude that the utilitarian cost-benefit approach comprises the preferred, or at least dominant methodology for analyzing the responsibility of researchers engaged in automated collection of publicly accessible data on-line, on-line data.

But the material we have reviewed thus far may be quite differently characterized as pointing toward a quite different approach that may be described as deontological in nature. From a deontological perspective, the *right* of a human being to informed consent might be held to be absolute – no matter how low the risks of harm may be. From this perspective, the right to informed consent cannot be overridden – especially from a cost-benefit perspective that would try to argue that the benefits of so doing so outweigh possible costs and risks to the individual. From the deontological viewpoint, such cost-benefit arguments, as framed by the belief that "the good of the many outweighs the good of the few," thereby runs the risk of becoming a moral slippery slope that can quickly justify wholesale violation of human rights if such violation (against a few) would provide sufficient benefit (for the many).

Courts have generally been willing to assume that unwanted electronic contact constitutes harm to a networked computer, creating an almost absolute right of exclusion against unwelcome file requests sent to otherwise publicly accessible computers – what might be characterized as a distinctly *deontological* claim directly challenging the utilitarian calculus suggested above -- namely, the webserver owner's *property rights*, including the right to prohibit a researcher's 'bot from accessing the webserver and its data, where such access is construed as a form of trespass against the webserver as chattel. By the same token, to maintain an archive of web pages for the sake of research, if not authorized by the web page owners, could be construed as copyright infringement. At least in the U.S. context, this concern is perhaps secondary, insofar as such archives might be arguable under the fair use provisions of the copyright law. Either way, of course, we seem to be left with potentially strong *deontological* claims: these rights cannot be overridden by the possible benefits of research – no matter how great those benefits might be. In short, the server owner appears to enjoy the moral equivalent of the right to informed consent. This conclusion raises significant questions as to whether current legal precedent, together with practice of deploying the standard for robot exclusion in fact amounts to a deontological norm regarding web site data.

The Morality of Law

As an initial matter, we note several caveats as to whether an emerging legal practice, either deontological or utilitarian, can inform the ethical obligations of researchers. This consideration implicates a long and unresolved debate in jurisprudence as to whether law should be obeyed because it has independent moral force – that is, that the creation of law creates a moral imperative -- or whether law should be obeyed because it reflects moral consensus, especially in democracies where law at least in theory constitutes the consensus of society (Hart, 1994; Fuller, 1969). To some extent, the question may turn to some extent on whether a particular law is *malum prohibitum* – for example, a regulatory rule for the convenience of the state, such as the 55 mile an hour speed limit, for which there is no reason to expect that a morally autonomous individual would inherently know this is the expected behavior – or *malum in se* – in an inherently moral rule, such as do not commit murder, that all morally autonomous individuals would be expected to know

is the expected behavior. One can advance a range of both natural law and utilitarian theories to support either view. But under either view, there is agreement that at least some of the time legal and ethical behavior will be correspond to one another.

At the same time, it seems equally clear that an individual's ethical and legal obligations may not be coterminous. In some instances, an individual's ethical obligations may be seen to exceed his or her legal obligations. To take a famous example, under the Anglo-American common law of tort, there is in general no legal duty to rescue – if an individual sees someone else drowning face down in a puddle of water, and could save them at absolutely no risk and essentially no inconvenience to himself, he has absolutely no legal obligation to do so -- but almost certainly has a moral obligation to do so, and would rightly be considered morally degenerate if he did not.

In other cases, one's legal obligations may be seen to exceed her ethical obligations -- for example, in the area of research ethics, it seems likely that many IRB requirements for human subjects research arise out of a desire to avoid legal liability rather than out of any requirement of respect for research subject's autonomy and welfare. At the extreme, it may in some instances even be necessary for an individual to engage in ethical acts that the law prohibits: hiding Jews from the Nazis, assisting escaped slaves on the Underground Railroad, or refusing to reveal subpoenaed sources of news information, for example. However, in general, many actions that depart radically from legal prohibitions will tend to be ethically questionable -- e.g., stalking or shooting abortionists in the belief that it is necessary in order to save fetuses from destruction.

In many instances, disparities between legal and ethical behavior will arise from differing measures or conceptions of harm, where "harm" is defined very broadly. In some cases, harm will involve an intrusion on the rights of the individual, regardless of the general effect on the good of society, perhaps even in spite of an effect on the good of society – a largely deontological claim. In other cases, the harm at issue may accrue to society as a whole, perhaps all at once, or incrementally as repeated harms to the individual. Deontological, rights-based claims may accrue out of instrumentalist motivations, as essentially *per se* categories of cost-benefit conclusions. Or, deontological claims may be weighed and balanced against other deontological claims, as

incompatible rights come into conflict. This continuum of analyses, from absolute deontological claims to absolute utilitarian claims indicates that law, like much of ethics, comprises an unusual admixture of utilitarian and deontological analysis. And, where one applies the ethics of deontology, but the law has chosen that of utilitarianism, or where one applies the ethics of utilitarianism when the law has chosen that of deontology, disparities occur.

In some instances, legal standards may give us an indication of the type of activity society considers harmful or not harmful. For example, trespass to land has long been held to constitute a harm *per se*; that is, the violation of an individual's control over land is a harm in itself, whether or not any actual physical damage was done to the property. To date, courts have tended to treat web sites much as they have traditionally treated land, being willing to assume harm, or to deter speculative potential harm, from unwanted electronic contact. This effectively makes unwanted electronic contact a harm in itself, without necessity of proving actual physical or financial harm. One reading of this line of cases is that society is willing to treat unwanted electronic contact as a *per se* category of harm that deserves ethical deference.

At the same time, legal standards may indicate categories of harm against which society refuses to recognize or offer protection. Web site owners might be "harmed," in the sense of receiving no direct benefit for their efforts, from the contact by researchers' data collection robots. But it is not quite correct to assert that automated research data retrieval constitutes free riding on the efforts or publicly accessible resources of web site creators. True, the web site owners will not get new business from the data retrieval as they might if the contact came from potential customers, nor will they get increased access and visibility as they might if the contact came from search engine or other indexing spiders' resource discovery robots. The web site owners do benefit in some measure, along with the rest of society, from the generation of new knowledge and understanding. Indeed occasionally they may benefit very directly from research that generates results that enhances their understanding of consumer behavior, web marketing, or business strategies. The overall societal benefit of the research, in which the web site owners participate, might lead us to conclude that they have not been harmed at all.

From a different conception of harm, web site owners might be "harmed" by the research contact to the extent that the published research also benefits their competitors, or reveals unsavory or inept practices in which they may be engaged. But it is unclear whether this latter type of harm constitutes harm that we should take into account in determining whether to refrain from conducting the research – damage to an undeserved reputation, or deterrence to an ill-advised activity, while constituting subjective harm from the web site owner's point of view, may not be actual "harm" of the sort recognized by society at all. Society may prefer that poor business practices be exposed, even though such exposure prevents a web site owner from pursuing his preferred course of action. Stated differently, it is not clear that there is a strong right to exclude researcher scrutiny in order to continue to engage in fraud or waste – whatever deontological claim the web owners might have does not extend so far, or is outweighed by the harms they are inflicting upon society and upon themselves.

The legal allowance for certain harms in the greater public interest, however, raises the question as to whether we are using individuals as a means to an end, rather than as an end in themselves, to the extent that we are regarding or disregarding their labor and creativity for broader social purposes. It is possible that it may be legally permissible, but still unethical, to appropriate certain types of information to the detriment of another. In other research contexts, it is likely that there is personal information that could be legally gathered, as it is subject to no recognized privacy or proprietary interest, but that gathering the information would disregard the research subject's autonomy or personhood such that the research would constitute an unethical practice. A parallel situation could arise with regard to automated web data retrieval, although the method of gathering, rather than the nature of the data is more likely to raise ethical issues.

Framing the "Meta-Ethical" Choice

Bearing in mind the caveats we have indicated with regard to the moral content and implications of law, and noting particularly the legal admixture of utilitarian and deontological justifications for particular policies, we are better positioned to consider the similar utilitarian-deontological mixture of utilitarian and deontological in ethical

research practice. To note that law and ethics mix deontology with their prevailing utilitarianism is by no means a critique. On the contrary, as the criticisms of utilitarianism noted above suggest, most ethicists argue that some combination of *both* utilitarianism and deontology is required for a more complete and robust ethical system. That is: the strengths of deontology may compensate for the deficits of consequentialism and vice-versa.

By noting this basic contrast in our moral thinking, we can often avoid unnecessary confusion – and, in some cases, resolve otherwise apparently irresolvable conflicts. In the case of automated data retrieval practice this conflict is not, as it may first appear, a conflict that pits utilitarian arguments against more or less equal arguments for a contrary view. Rather, the conflict here is more fundamental – namely, between a consistently utilitarian approach and basic deontological rights surrounding notions of property and express intentions. From this larger perspective, the two sets of arguments thus run the risk of simply being irrelevant to one another, insofar as each is grounded on distinct, and perhaps incommensurable starting points. More positively, however, this perspective suggests that the appearance of a commensurable conflict is just that – an appearance. By viewing the information policy arguments as consistently utilitarian, vis-à-vis the countervailing positions as deontological, the debate shifts to a second "meta-level" – one that forces us to ask: of these two ethical approaches *per se*, which *should* supersede the other in case of conflict?

The Question of Consent

The "meta-ethical" choice between utilitarian and deontological criteria is well illustrated by considering the questions of consent, implied consent, and informed consent in regard to automated data retrieval. Much of the legal analysis surrounding the trespass to computers cases hinges upon issues of consent: trespass occurs only if contact with the computer is unauthorized, and courts have been willing to infer some degree of authorization from the fact that the computers are connected to a publicly accessible network. Clearly the computer owners desire or permit some degree of contact with their computers, otherwise they could disconnect the machines from the network, or use password protect protection on their web sites, or take other action to restrict access.

However, such implied consent can be expressly revoked. Courts have found evidence of such revocation in the terms of service posted on some web sites, stating that certain types of access by bots is prohibited. The posting of "robot.txt" files, machine-readable files implementing the "standard for robot exclusion," instructing bots to refrain from crawling all or part of a web site, has also been taken as evidence of revocation.

To be sure, neither of these indicators of revocation may ever be actually seen by the individual deploying a software robot. Terms of service pages may be buried within web sites beneath several layers of linked pages; robot.txt files are not meant to be seen by a human at all, and do not necessarily contain any human-readable text are meaningless to individuals not already familiar with the standard. In such cases, the notice of revocation may be constructive rather than actual; it may be that the owner of the robot should have known of the explicit revocation of implied permission rather than did know of the explicit revocation of implied permission. Such a rule of constructive notice effectively shifts the burden of determination, placing the onus of investigating whether the robotic contact is permissible to the individuals deploying robots, rather than laying the burden of notification on the web site owner.

This places a significant burden on the researcher employing software data collection bots; since copyright liability is strict, and a number of courts have treated liability for trespass to computers as similarly strict, the researcher may bear the full burden of examining the web site for signs of revocation. In the case of the robot.txt file, the indicator of explicit revocation may be automated, but software bots cannot read or comprehend human-readable terms of service postings. If the terms of service or robots.txt file indicate that robot crawling is not permitted on the web site, the researcher may then bear the additional burden of seeking out and contacting a human agent to obtain explicit consent for the date retrieval.

Requesting such consent of course entails costs and burdens on the researcher – including the risk of being rejected, thus seriously undermining one's research from the outset. The trespass cases reviewed here include instances in which requests for permission to crawl a web site were explicitly denied, but commercial crawlers determined to proceed anyway, reasoning that their activities as a legal matter required no

permission, as they would accrue to the public benefit. In the commercial context, courts have not tended to look kindly on such self-help in the face of explicit denials of permission. Researchers who do not seek consent from website owners similarly run the risk of knowingly violating the web site owner's wishes and property rights. And an additional methodological concern is that requesting consent may pollute the data, for example causing the site operator to alter or mask information that may reflect poorly on the transactions or business model attending the site.

This calculus of benefits and detriments is characteristic of the utilitarian approach, and the factors listed above suggest that such a cost-benefit analysis may proceed at different levels. At one level, the level of public policy, the general benefit to society from might be compared to the detriment of the web site owner. Under this approach, the value of the research may appear to exceed the relatively small detriment to a given web site owner. Obtaining consent manually is laborious, negating the advantage of the automated research technology, and potentially deterring a good deal of beneficial research. As in the *Feist* case or in copyright fair use, a public policy cost/benefit analysis might well permit unauthorized automated research activity in the absence of consent due to the overall public benefit – but again, this may be at the cost of using web site owners or their property as a means to general public benefit.

But at a different level, that of an individual researcher, a cost-benefit analysis may yield quite a different result. That is, to put the point on it: researchers may choose to respect the rights of web server owners – not because those rights are paramount (the deontological argument), but because from a utilitarian, cost-benefit approach, the potential costs of violating those rights might be greater than a researcher would prudently risk. A risk-averse conclusion at the level of individual calculation may treat the rights of the web site owner as a fait accompli, to the detriment of the general good of society, and ironically, in essential alignment with the outcome under a deontological view of the rights of the web site owner. In addition, there is a strong thread of ethical "good Samaritanism" in Internet research ethics – i.e., instances in which researchers go above and beyond the minimal requirements of given ethical codes and extant laws, in order, for example, to respect what they see as important expectations concerning privacy, autonomy, etc. – even though doing so issues in significant costs as this

complicates their research efforts, and, in some instances, may even run the risk of canceling the research project entirely.[2]

The Informed Consent Model

In a different context, Amy Bruckman and James Hudson have argued that such costs are part of a larger set of reasons for *exempting* researchers from requesting informed consent with regard to at least certain forms of online research, such as chatrooms (2004). As noted above, this is characteristic of U.S.-based approaches to research ethics are typically, though not exclusively, utilitarian. So, in the U.S., exemptions to an otherwise primary obligation, such as seeking informed consent, are often justified in utilitarian, cost-benefit terms. By contrast, such exemptions are much harder to justify in the European and Scandinavian contexts. On the contrary, the strongly *deontological* cast of research ethics – most notably, in Norway (National Committee for Research Ethics in the Social Sciences and the Humanities [NESH], 2001, 2003) – emphasize first of all a range of Human Subjects protections that must be satisfied for research to be undertaken. These protections are absolute in the sense that they cannot be overridden by an argument that the possible benefits of any research that would violate these protections would be sufficiently significant to the majority of society.

To be sure, on both sides of the Atlantic, social scientists recognize the problem that adequately *informing* a subject of the purposes of a research project may change the subject's behavior in ways that invalidate the study. And for this reason, modified informed consent as well as exemptions under some circumstances are judged as ethically allowable – for example, an initial, but incomplete – possibly deceptive – description of the experiment, followed by a more complete disclosure at the conclusion of the experiment, with the subject having the option of refusing to allow his/her data included at that stage.

Such debates track several of the issues identified in relation to automated data retrieval, and the general issue of consent appear to be an attractive point of congruence

---

[2] "Good Samaritan ethics" is a notion introduced by Judith Jarvis Thomson in her landmark article (1971). For an example of how one researcher decided to stop her research in order to contact a young woman who had left information on her homepage that could have resulted in direct harm, see Løfberg (2003). For further discussion, see especially Stern (2003, 2004).

between legal and ethical obligations. Under Western ethical practice, we typically consider consent of the subject to be both a cure for imposition of harms or risks in research and as a validation of the research subject's autonomy; assuming that the subject is competent to make such choices, we respect the subject's right to assume or decline research risks. This closely follows Western models of contractual assent, which assumes that a competent individual is in the best position to choose or decline legally binding obligations -- indeed, the same consent form typically serves to deal with both ethical and legal research obligations. Assent to automated data retrieval from web sites would presumably cure both ethical and legal objections to contact by data collection bots.

The analogy to informed consent has been previously analyzed in regard to on-line technical practices, including some that implicate the issue of trespass. Millet et al. (2001) have argued that informed consent principles should apply to the placement of "cookie" files on the computers of users who access web sites where cookie technology is used to identify users and their preferences, and have analyzed the compliance of successive browser versions with such principles. Although unconsented cookie placement could be deemed a form of trespass, (Siebecker, 2003) the major concern of these analyses is the potential for such cookie deployment to invade privacy by creating user profiles and track user activities across co-operating websites. Specifically, these studies recommend use of informed consent in order to address the otherwise unconsented gathering of personal, personally identifiable information that cookie technology could collect without giving web browser users the opportunity to understand what information was collected and how it was to be used, as well as the opportunity to accept or decline to participate in such information collection.

Differentiating Automated Data Retrieval

The type of automated data gathering using bots considered here may be said to differ markedly from the cookie situation, in both technical and ethical characteristics. Unlike cookie placement, which deposits a file on an individual computer to mark computer personal and personally identifiable activity that would be otherwise unobserved, automated software data collection bots deployed for research request from another computer information that is otherwise publicly accessible, and typically not personally

identifiable. Thus no alteration is made to the state of the code on the crawled server, and the danger of personal information collection is greatly diminished if not altogether absent. While the individual and personally identifiable nature of cookie activity lends itself to an application of informed consent, it is not at all clear that such an extension of informed consent is appropriate in the use of research 'bots. The public and impersonal nature of research bot activity means that no credible threat to the autonomy or personal integrity of an individual is present in the case of automated data retrieval.

Indeed, it may make sense to distinguish between treating a web site owner (or any individual) as a means to an end, as opposed to treating their *property* as a means to an end – which presumably it is, even for them. As we have noted, property, whether tangible chattel or intellectual property are viewed in the U.S. as largely as instrumental creations, developed for utilitarian purposes in the first instance. Some Hegellian models of property suggest that some certain types of property may have a basis in the individual's personhood, where the property is integral to their individuality or personality – a wedding ring, for example. (Radin, 1982) But even under a Hegellian model, it seems somewhat far-fetched, to argue that web pricing data is strongly bound up with someone's personal identity. Absent a strong "personality" theory to animate a deontological approach, an approach that balances property interests against the public good may seem more attractive.

Moreover, given that many or most of the web sites that will be the subject of automated data retrieval are corporate in nature, the question of personhood in turn raises the question as to whether corporations, as juridical "persons" deserve the degree of respect that we would accord to natural persons. Do we need to worry about using corporations as means to an end? The law typically treats corporations as equivalent to natural persons for a variety of limited purposes, i.e., they have the right to own property, the "right" to be criminally liable and punished via fines (which implies moral autonomy, in order to be punished for improper choices), at least a limited right to engage in free speech under the First Amendment, etc. On the other hand, they are treated as persons for only limited purposes, and do not share many of the other legal recognitions of natural personhood, i.e., they have no right to vote, no right against self-incrimination, no right to

counsel, no right to basic social services, etc. Consequently, the legal signals are at best mixed as to the extent to which corporate "personhood" deserves moral respect.

Given that corporations are by definition entirely instrumental, created as means to an end – for the management and increase of shareholder investments – it is unclear how we can consider them as ends in themselves. It may be that, again instrumentally, we should treat them or their activity with some degree of respect as a means of respecting the status of the individuals who comprise the corporation. There is some precedent for this in the area of American First Amendment jurisprudence, where the corporation's right to speak is largely derivative of the interests of the underlying group of shareholders. But in such cases the corporation serves merely as an instrument for expressing the rights of natural persons, so that a strong identification of personhood with corporate property seems unlikely, suggesting that both the corporate entity and corporate property might properly be treated on a utilitarian basis.

Conclusion

In this case, by recasting the ethical obligations of the researchers within this larger framework, we hope to bring into sharp focus the contrast between certain aspects of the law as utilitarian, on the one hand, and, on the other hand, different and contrasting aspects of the emerging law regarding servers as chattel, the "robot.txt" specification and its affiliated practices as deontological. Stating the point in this way, we hope, makes clear, that the conflict between the U.S. practice of exempting some research from the requirement of informed consent when risks are low and costs are high, and a European, especially Scandinavian insistence on informed consent as one of several human subjects protections, no matter the cost, is a conflict at a first level only. If left at this level, the conflict would remain an irresolvable either/or: either the U.S. is right or the Scandinavian countries are right – but not both, so that one would be forced to choose between them.

By contrast, at a second "meta-level" – this conflict can be resolved in an interesting way: As we have suggested, when viewed at this more fruitful "meta-level," substantial questions exist as to the applicability of deontological approaches outside the milieu of human subjects research, in an environment of automated interaction between

software bots. Claims to personhood - and thus to deontological rights of a near-absolute sort - on the part of corporations is not fully persuasive; rather, the status of corporations as persons is only partial - and this for utilitarian rather than deontological reasons. This suggests in turn that the more utilitarian considerations regarding benefits of research should come into play at a first level: utilitarian research benefits outweigh any rights corporations may have, at least as "persons." Most importantly they may outweigh rights to informed consent, because, second, in this instance, at a meta-level the utilitarian considerations outweigh the deontological claims.

If such an analysis leads us to accept the primacy of the utilitarian approach, this in turn suggests that by taking up this framework, we indeed move beyond conflicts on a first level that might otherwise seem intractable. In an important sense, there is no conflict, because each ethical conclusion follows from different ethical premises, either utilitarianism or deontology. By moving to the meta-level, as we hoped, we are able to argue for a resolution to the first-level conflict and particular issue raised by automated data gathering techniques - a resolution that should be fruitful insofar as it offers researchers ethical guidance on the use robots without asking for informed consent while also complying with the broader policy dictates of the law.

References

Adam, A. (in press). Ethics for Things. *Ethics and Information Technology*.

Allen, G. & March, S. (2000). Developing Internet agents: A tutorial using Visual Basic 6.0. In W. J. Orlikowski, P. Weill, S. Ang, H. C. Krcmar, & J. I. DeGross (Eds.), *The Proceedings of the 21st International Conference on Information Systems* (Dec. 10-13, Brisbane, Australia) (pp. 733-738). Atlanta: Association for Information Systems.

Baye, M. & Morgan, J. (2001). Information gatekeepers on the Internet and the competitiveness of homogeneous product markets. *American Economic Review*, *91*(3), 454-474.

Burk, D. L. (2006). Privacy and property in the global datasphere. In S. Hongladarom & C. Ess, (Eds.), *Information technology ethics: Cultural perspectives* (pp. 94-107). Hershey, PA: Idea Group.

Burk, D. L. (2000). The trouble with trespass. *Journal of Small & Emerging Business Law*, *4*, 27-56.

Bruckman, A. & Hudson, J. (2004). Go away: participant objections to being studied and the ethics of chatroom research, *The Information Society, 20*(2), 127-139.

Clay, K., Krishnan R., & Wolff, E. (2001). Prices and price dispersion on the web: Evidence from the online book industry. *The Journal of Industrial Economics*, *49*(4), 521-539.

Clemons, E. K., Hann, I., & Hitt, L. M. (2002). Price dispersion and differentiation in online travel: An empirical investigation. *Management Science*, *48*(4), 534-549.

eBay Inc. v. Bidder's Edge, Inc., 100 F.Supp.2d 1058 (N.D. Cal. 2000).

Elkin-Koren, N. (2001). Let the crawlers crawl: On virtual gatekeepers and the right to exclude indexing. *University of Dayton Law Review*, *26*, 179-209.

Epstein, R. (2003). Cybertrespass. *University of Chicago Law Review*, *70*, 73-88.

Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., et. al. (1999). *Hypertext Transfer Protocol -- HTTP/1.1*. Retrieved September 9, 2003, from http://www.w3.org/Protocols/rfc2616/rfc2616.html.

Friedman, B., Howe., D., & Felten, E. (2002). Informed consent in the Mozilla browser: Implementing value sensitive design. In R. H. Sprague (Ed.), *Proceedings of the 35th Annual Hawaii International Conference on System Sciences, Vol. 8.* (p. 247). Los Alamitos CA: IEEE Computer Society.

Fuller, L. L. (1969). *The morality of law*. New Haven, CT: Yale University Press.

Grodzinsky, F., Miller, K., and Wolf, M.J. (in press). The Ethics of Designing Artificial Agents. *Ethics and Information Technology*.

Hahn, J. (2001). The dynamics of mass online marketplaces: A case study of an online auction. In J. Jacko & A. Sears (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (March 31-April 5, Seattle, WA) (pp. 317-324). New York: ACM Press.

Hart, H. L. A. (1994). *The concept of law* (2nd ed.). Oxford: Oxford University Press.

Johnson, D. and Miller, K. (in press). Un-Making Artificial Moral Agents. *Ethics and Information Technology*.

Kauffman, R., March, S., & Wood, C. (2000). Mapping out design aspects for data-collecting agents. *International Journal of Intelligent Systems in Accounting, Finance, and Management*, *9*(4), 217-236.

Kaufman, R., and Wood, C. (2000). In W. J. Orlikowski, P. Weill, S. Ang, H. C. Krcmar, & J. I. DeGross (Eds.), *The Proceedings of the 21st International Conference on Information Systems* (Dec. 10-13, Brisbane, Australia) (pp. 145-151). Atlanta: Association for Information Systems.

Kostner, M. (1994). *A Standard for Robot Exclusion*. Retrieved September 9, 2003, from http://www.robotstxt.org/wc/norobots.html.

Kostner, M. (1995). *Robots in the Web: Threat or treat*. Retrieved December 21, 2007, from http://www.robotstxt.org/wc/threat-or-treat.html.

Kostner, M. (1996). Evaluation of the Standard for Robots Exclusion. Retrieved December 21, 2007, from http://www.robotstxt.org/wc/eval.html.

Löfberg, C. (2003). Ethical and methodological dilemmas in research with/on children and youths on the Net. In M. Thorseth (Ed.), *Applied ethics in Internet research* (pp. 141-154). Trondheim, Norway: Programme for Applied Ethics, Norwegian University of Science and Technology.

McGowan, D. (2004). Website access: The case for consent. *Loyola University of Chicago Law Journal, 35*, 341-386.

Millet, L. I., Friedman, B., and Felten, E. (2001). Cookies and web browser design:Toward realizing informed consent online. In J. Jacko & A. Sears (Eds.), *Proceedings of the SIGCHI   Conference on Human Factors in Computing Systems* (March 31-April 5, Seattle, WA) (pp. 46-52). New York: ACM Press.

NESH (National Committee for Research Ethics in the Social Sciences and the Humanities). (2001). *Guidelines for research ethics in the social sciences, law and the humanities*. Retrieved December 21, 2007, from http://www.etikkom.no/NESH/guidelines.htm.

NESH (National Committee for Research Ethics in the Social Sciences and the Humanities). (2003). *Research ethics guidelines for internet research*. Retrieved December 21, 2007, from http://www.etikkom.no/Engelsk/Publications/internet03/view.

O'Rourke, M. A. (2000). Shaping competition on the Internet: Who owns product and pricing information? *Vanderbilt Law Review*, *53*, 1965-2006.

Pan, X., Ratchford, B. T., & Shankar, V. (2002). Can price dispersion in online markets be explained by differences in e-tailer service qualities? *Journal of Academy of Marketing Science*, *30*(4), 433-445.

Radin, M. J. (1982). Property and personhood. *Stanford Law Review*, *34*, 97-1015.

Register.com, Inc. v. Verio, Inc. 126 F. Supp. 2d 238 (S.D.N.Y. 2000).

Shankar, V. Ratchford, B. T., & Pan, X. (2002). Equilibrium e-tailer prices: Pure play vs. bricks and clicks e-tailers. Paper presented at the *2002 Marketing Science Conference*, Edmonton, Canada: University of Alberta.

Sheng, Y., Mykytyn, P., and Litecky, C. (2005). Competitor Analysis and its Defense in the E-marketplace. *Communications of the ACM*, *48*(4), 107-112.

Siebecker, M.R. (2003). Cookies and the common law: Are Internet advertisers trespassing on our computers? *Southern California Law Review*, *76*, 893-952.

Stern, S. R. (2003). Encountering distressing information in online research: A consideration of legal and ethical responsibilities. *New Media & Society, 5*(2), 249-266.

Stern, S. R. (2004). Studying adolescents online: A consideration of ethical issues. In E. Buchanan, (Ed.), Readings in virtual research ethics: Issues and controversies (pp. 274-287). Hershey, PA: Information Science.

Thomson, J. (1971). A Defense of Abortion. *Philosophy & Public Affairs, 1*, 47-66.

Data as Representation: Beyond Anonymity In E-Research Ethics[1]

A. Carusi
Oxford e-Research Centre
7 Keble Road
Oxford
OX1 3QG, UK
Oxford University
annamaria.carusi@oerc.ox.ac.uk

Abstract

As e-research practices begin to be adopted in medical and social science research, a number of ethical challenges are being encountered, and doubtless there are several hidden ethical implications which will emerge as e-science matures. Concerns so far have centred on anonymity, confidentiality and privacy. This article proposes a representational framework for thinking about data in e-research which takes it beyond this concern. It proposes five representational models which can be used to probe the ethical issues around data derived from human subjects, and as an ingredient in ethical decision-making in this domain.

Introduction: E-Research

There are currently initiatives in the UK, USA and Europe to extend to medical research and health care, as well as sociological and anthropological research, the tools and technologies of so called 'e-science' or e-research.[2] This shift has brought with it ethical challenges which researchers, institutions and ethicists are hard-pressed to meet. The ethical challenges are generally placed under the rubric of privacy, confidentiality and anonymity. However, underlying these is a dominant conception of data as

---

[1] Research for this paper was supported by ESRC grant RES-149-25-1022.
[2] See for example www.nesc.ac.uk and http://www.eu-egee.org/.

information. This article proposes an alternative conception of data as representation which widens the scope of ethical considerations to be taken account of 1) in the design of the tools and technologies for e-research that deals with human subjects, 2) in the ethical guidelines for researchers and institutions, and 3) in the general understanding of ethical aspects of research of this type and responsibilities of both researchers and research participants.

e-Research was originally thought of as science carried out by means of high performance computing which required a 'Grid'. The Grid was initially envisaged as a computational grid, that is, distributed sets of computers co-operating on a calculation. Currently, there are shifts away from a grid as such, and reliance on other Internet technologies (such as 'cloud' technologies), and the infrastructural aspects of this form of research are focused upon. Above all, e-research technologies are thoroughly social and institutional as they relate to access to research resources. As such, the way in which these relations are understood is crucial to the design of e-research tools and technologies.

The e-research applications that are the focus of this paper are those which make use of one or more of the following:

- Large-scale databases (video, audio, image, survey questionnaire responses, biographical, demographical and health information, etc.)
- Resources for data-sharing, either for data-mining, for re-use of data, or for collaborative analysis and interpretation of data.
- High performance computing or computational resources such as metrics, modelling, simulation and visualisation.

Whereas in the initial natural and physical science applications the most important social implications are to be found in the sphere of co-operation and collaboration, and the ethical implications are primarily in the domain of trust among researchers in their institutional settings, the broadening of e-science from natural science to health and social sciences has brought with it the entire gamut of social relations between researchers and others who manage and deal with data, and the human subjects from whom data are obtained. In the UK, e-social science projects are funded and promoted in various ways, with a major initiative in the UK being the National Centre for e-Social

Science, under which there are 7 large scale projects, one 12 small scale projects (see www.ncess.ac.uk for a description of these projects). It is the NCeSS projects in particular that have formed the background of this paper. Other important examples are to be found in anthropology (for example, an extensive digitisation project undertaken by the Cambridge University Museum of Anthropology and Archaeology, and several other examples on the Internet) and the quantitative data base of the British Household Panel Survey (Carlson et al., 2006). However, current digitisation drives by funding agencies who make it a condition of funding that data are archived in a data base, and by public health services which are digitising and archiving health records will result in massive amounts of information about people being grid or Internet-enabled, with the attendant issues, problems and concerns relating to anonymity, confidentiality and privacy (Carusi & Jirotka, 2007).

Among the lessons learned from projects in the medical area, such as e-Diamond, a project to build a federated database of mammograms for distributed readings and training (Jirotka et al., 2005), is that legal and ethical restrictions on what can be done with data obtained from human subjects could severely restrict e-research applications of health and social science (Hinds et al., 2007). While the only restrictions on physicists' sharing and re-use of data are their own collaborative practices, there are currently strict limitations on the sharing and re-use of data obtained from human subjects, and even where there are not, there may still be ethical reasons that preclude this. However, the effect of legal and ethical restrictions and personal ethical reservations of some researchers will not be to shut down health and social science applications of e-science, but rather to shift them towards methodologies which are more amenable to these cyber-infrastructural environments.

We should expect to see an increasing mix of quantitative and qualitative methodologies, with images, audio and video which are often used for qualitative social science research, being subjected to increasing quantitative measurement. Data merging and mining are being increasingly used, as are other computational resources such as webmetrics for social network analysis. There are extremely important shifts in the mode of doing health and social science unfolding at the present time, and with them will come extremely important social, ethical and political shifts too. It is important to keep as rich

a variety as possible of different types of research and different methodologies alive and viable on the Internet and on specialised research grids. Important, that is, both for a rich and nuanced understanding of human beings and societies and for the social and political function of those understandings. In light of this, it may be necessary to re-think some of the ethical principles that have informed research practices, or to re-think the relationship between researchers and subjects of research.

These developments are currently unfolding and as yet cannot be studied with the cool detachment of hindsight. Instead, some foresight is needed to try to project what hidden implications these tools and technologies may have in order to take these into account for both the design and development of the technologies and the way they are socially and institutionally embedded. Indeed, this has already been sharply felt in Internet research ethics. In order to try to address these challenges, the Association of Internet Researchers set up a committee, led by Charles Ess, to offer guidance to researchers in facing a very different ethical landscape (Ess et al., 2002). E-research differs in important respects from Internet research, as it does not (necessarily) carry out research on subjects already on the Internet with the purpose of understanding Internet behaviour. Social science use of e-research technologies is different in that the research is mostly carried out on subjects in the offline world; it does not study subjects on the Internet, but uses cyber-infrastructural resources for the archiving, analysis or interpretation of data. However, there is a need for a similar process of consultation and ongoing debate in order to try to meet the ethical challenges attendant on the new technologies. The aim of this article is to broaden the range of factors that are traditionally considered to fall under the rubric of research ethics, beyond the central three issues of anonymity, confidentiality and privacy, to more general representational aspects of data.

The article begins with a consideration of notions of identity in research relating to human subjects and defends the ethical significance of thick identity. In the second section, thick identity is in turn related to a view of data as a form of representation as well as only information, via the notion of 'representation as'. In the third section, I consider the possible effects of digitisation on data, the ways in which it can be processed which are relevant to an understanding of data as representation. In the fourth section, an

outline of five representational models is given as frameworks for understanding different forms of data and the relation they have to researchers and research subjects who are implicated in them. A wider study would consider institutional aspects as well, but this cannot be covered here.

Identity

While not in any way diminishing the importance of the problems of anonymity, confidentiality and privacy, this paper suggests that there are ethical issues and concerns beyond these, which emerge when data are regarded as representation rather than as information. Concerns around anonymity, confidentiality and privacy revolve around the possibility of re-identifying individuals, and the information that can be obtained about re-identifiable individuals. I distinguish between two senses of identity: thin and thick. The distinction can be seen from the point of view of what is believed or known on the basis of data. Thin identity is the identity of a particular individual as a re-identifiable entity. Proper names pick out particular individuals and have to do with thin identity. Thick identity is a matter of that individual's experience of their own personhood, their own subjective or psychological sense of who they are.

I wish to suggest that thick identity can be a matter of ethical concern even when it has been detached from thin identity. Thus even if anonymity is guaranteed, there are ways of treating thick identity that it would be appropriate to make an ethical judgment about. One aspect of thick identity is the narratives that people use to make sense of their lives and circumstances. Taking over or appropriating images, quotations, stories or other ways in which thick identity is mediated, possibly for some purpose to which consent has not been given, is an instance of a use of anonymised data which could be considered as apt for ethical judgment.[3] Of course, we are accustomed to this happening, for example in journalism and in biographies, but firstly, the fact that we are does not mean that they can therefore no longer be considered as cases for ethical judgment; and secondly, it is important to recall that we are dealing with research cases, and the type of agreement or contract that there exists between researchers and their subjects is not the same as that

---

[3] See Tomaselli (1996) for an interesting treatment of this topic. For narrative discourse and identity, see Bamberg (2004), and for an online development of this view, and the ethical guidelines given for it, see www.talkinglongterm.co.uk.

which exist between journalists and their interviewees. Considering this kind of appropriation as something of moral relevance has to do with the representational content of the data, rather than with information relating to an individual.

Another way in which thick identity is experienced is by means of identification with a group, or a collective identity. Our current social and political climate means that head scarves are as much a matter of the individual identity of the women who wear them, and of their group identity. The head scarf is a metonymic signifier of something that goes beyond allegiance to the group, but of an experience of personhood in terms of the group. The ethical questions over how groups are portrayed are now familiar from decades of cultural studies into gender, race, and homosexuality. Thus, even if no particular individual is identified by some way of handling data, some aspect of their thick identity – as carried by data of some form – can be used to construct an identity for a group and this construction has ethical implications.

Information and Representation

Data are normally seen as information rather than representation. However, thick identity is often a matter of representation rather than information. For example, the sense of self is often expressed and formed in a kind of personal narrative, a form of representation with great psychological power. Images and metaphors are important in this too. The distinction can be put as follows: the thin identity of an individual has to do with the relation between that particular individual and the fact that they have a certain medical condition; the thick identity of that individual has to do with his or her representation as a victim of that condition, rather than as a fighter, or survivor, or simply neutrally as having it. This is where the notion of 'representation as' as it is theorised in philosophy of art and cultural and critical studies has potential to be useful as a way of understanding fundamental ethical concerns. 'Representation as' derives from the insight that there is not a single or unique way to represent an item or a person (Goodman, 1976; Gombrich, 1983). Given the multiplicity of ways in which people can be represented, the fact that they are represented in one way rather than another can be motivated by values, interests, prejudices which are not obvious either to those being represented nor to those doing the representing. Verbal and visual representations are a crucial arena for the formation and

playing out of personal and social values. The social values embedded in representations are often the hidden and implicit motivators of explicit moral judgments. Philosophers have different ways of distinguishing different aspects of the moral terrain, but one way to see this is in terms of Habermas's conception (drawn from phenomenology and hermeneutics) of the lived world as the context and horizon of expectations in which group solidarities are formed around certain values as a background for ethics proper (Habermas, 1988).

It is by considering data as representation that some of the ethical concerns around it that would otherwise remain hidden come to light. This attunement to the ethical dimension of data as representation has an affinity with disclosive ethics with respect to technologies (eg Brey, 2000): just as hidden and implicit values are embedded in the design of technologies, so are they embedded in data considered as a form of representation.

In this section I have argued that thick identity is closely related to 'representation as', that is, to the way in which subjects are represented, and that representation is value-laden in a way that makes it apt for ethical consideration. I will not try to broach the question in this article of what kind of ethical consideration or ethical action follows from this view of data as representation. The point I make is a small one: even if anonymised and not traceable to any particular individual, the representational character of data has ethical implications which need to be considered as we try to work out our overall reactions to e-research in the social and health sciences.

Possible effects on data of Digitalisation

Digitalisation processes allow for a myriad different possibilities of manipulation of data, from the very basic construction of data bases, to more intrusive manipulations (Gross, Katz, & Ruby, 2003). In this section I consider standardisation and reification, re-contextualisation, manipulation, and computationally assisted interpretation. This is not intended as an exhaustive list, though they do seem to be central to many e-science informed modes of research.

Standardisation and reification

In the case of e-social science, a particular set of issues emanate from the needs of standardisation. A shared database is not useful unless there are standardised ways of presenting and dealing with data. If data are to be made available in data bases for possible re-use or collaborative use, they must be standardized.[4] In addition, the computations that are accessed via the grid could require data that are delivered in a certain way: for example, in the case of a multi-modal analysis of natural language case, a computation to measure the range of the head nods that accompany language requires that the image be taken in a particular frame, from a particular distance and angle (Knight et al., 2006). How it is standardised makes an enormous difference. The choice to present a three dimensional object in two dimensional format; the choice to have all images of head nods framed in a particular way, the choice to cut and crop in a particular way, to zoom in on some aspect rather than on another, the choice to associate particular labels / metadata with images: these are all examples of digitalising processes.

Standardisation is connected with reification. Reification occurs when the aspect under which something is known comes to be identified with that thing. For example, the anthropologist Marilyn Strathern, describes it as follows:

> By reification I simply intend to point to the manner in which entities are made into objects when they are seen to assume a particular form ('gift', 'exchange'). This form in turn indicates the properties by which they are known and, in being rendered knowable or graspable through such properties, entities appear as "things." (Strathern, 1999, p. 13)

Standardisation feeds into reification because it forces all items in a data base to be represented under the same aspect which then defines by which properties the items represented are known. Reification of the subjects represented under those aspects is facilitated. Reification can sometimes result in data acquiring an independent status from subjects from whom it was derived. This has different ethical consequences depending on

---

[4] In addition, it is often required that the way in which data are standardised be transparent to other researchers. The provenance and history of data are extremely important for other researchers if they are to trust it. See Carlson & Anderson, 2006, on astrophysics, and Jirotka et al, 2005, on mammographies.

the type of data they are: for example, a discrete item of data (image, video, quotation) embedded in a larger data base, or a series of data all connected to the same person.

Re-contextualisation

De-contextualising: e-research data are made accessible to researchers beyond the initial data gatherers. They are shared either for collaborative research, or for re-use for other research.

Re-use of data means taking them out of their context of collection – that is, out of the situatedness from which they emerged. Since context plays a significant role in specifying the representational content of items, this would evidently have an impact on the content, meaning and significance of data. Using audio or text alongside images cannot but suggest that they are meant to illuminate each other, and should be interpreted accordingly. Any film maker knows that; however, the new digital media make these possibilities available to anyone with the technological means and also make it possible to take apart contextualising features of a representation. De- and re-contextualisation can also have ethical implications. Contextualising a photograph of a slender girl in a data base of sufferers from anorexia cannot but represent her as an anorexic.

It is instructive to use anthropology as an example, as museum collections have historically been less regulated than other forms of social science, and can be instructive as to what can happen in the absence of regulation. A random example is the Smithsonian Natural History Museum's online exhibition dedicated to 'African Voices', where various images of people and artefacts, as well as voices, can be accessed.[5] The site presents an interesting case for political and ethical analysis of the effects of digitalisation, where the hypermediated contextualisation offers itself to critique in terms of its implicit values regarding the representation of 'others'. The point here, however, is that digitalisation makes possible endless contextualisations and re-contextualisations, with not only the site presenting its own array of contextual possibilities but also making it possible for viewers to continue the process of de- and re-contextualisation. For

---

[5] http://www.mnh.si.edu/africanvoices/

example, it is possible to download any of the images on the site – in fact, sometimes the viewer is invited to do so: the image of a mudcloth, for example, can be downloaded as a screen saver. This is quite a radical de-contextualisation of these images, one which results in their representational content being quite drastically affected (one can't see it is a cloth at all, let alone a mudcloth; the shapes on it play out their significance among the shapes of the viewers life, and not those of the producer's (cf Geertz, 1976). What can be done with the relatively insignificant mudcloth can be done with any of the images or audio. The ethics of appropriating representations of others and of their world are debatable and I am not suggesting that the ethical implications of the new possibilities of interacting with data brought into being by digitalisation are necessarily negative and exploitative. My point here is that the digital medium makes these re-contextualisations possible, with the concomitant effects on the meaning of the data, and the associated ethical issues.

Manipulation

Standardisation and (re-)contextualisation in data bases are both facilitated by the greater possibilities for manipulation of digital data. This is particularly evident in images: these can be cropped, the colour can be changed, montage is far easier, and so on (Gross et al., 2003). It is in view of the possibilities of manipulation that it is important for researchers to have a full biography of data that they use or re-use in their own studies. It is not scientific fraud that is the main ethical worry here, but rather the type of manipulation that falls within the bounds of normal scientific procedure. Any of these manipulations for the purposes of standardisation could result in differences in the 'representation as' aspect of data.

Interpretation and computation

e-Research applications use digitalised data in order to be able to use on them analytical and interpretational resources that would otherwise be inaccessible or difficult. In the first place, collaborative interpretation is made possible by the e-science infrastructure. For example, this was a feature of e-Diamond, the grid-enabled infrastructure to allow mammographies to be read and interpreted at different sites of expertise. A standard kind

of interpretational resource is annotation; in e-science applications there are two possibilities that stand out: collaborative annotation and multimedia annotation (for example, running transcript and video or audio side by side and being able to annotate both at the same time). Annotations then become a part of the context of the data, and inflect its representational content. With these two possibilities come other ways in which 'representation as' can be affected: first of all by the intervention of other researchers besides those who gathered the data from subjects and have a direct relationship with them; second by the juxtaposition of different media.

Data-mining is another technique which is facilitated by digitalised data in large data bases. Data-mining can latch on to the annotations or other tags attached to data by researchers. Data-mining leads to re-contextualisation, and sometimes it can lead to a re-contextualisation which results in the data coming to be categorised under social groupings which the data subjects – nor indeed the researchers who dealt with them directly – could foresee (Tavani, 2004). The representational content of the data thus is inflected by a social group which is not consciously part of the data subject's thick identity. This does not mean that it not a part of thick identity: most people of the petit bourgeois class in Karl Marx's time would not have recognised themselves in the class structure as described by Marx, but this does not in and of itself invalidate what Marx had to say about their consciousness. But this in itself shows that these re-groupings are not value neutral.

Standardisation, reification, re-contextualisation, manipulation, data-mining, annotation all have both methodological and ethical implications. They can privilege or occlude some analytical and interpretational possibilities in specific studies, and lead to systematic overall analytical and interpretational trends in social science.

The fact that these technologies are not ethically neutral need not lead to dismay and attempts to curtail them. They could have positive as well as negative implications. For example, digitalisation can break previous norms of standardisation (something which art has always done) and disclose what these norms are, thereby de-mythologising them; the re-configurability of data collection means that different aspects and stories about the people represented can emerge, and the manipulability of digital images and

other information means that they can also be re-appropriated by their subjects. There is also a greater interactivity with data and data bases, and with digital collections, access allowing. These are powerful media, and because they are, the question who has access to them, and what sort of access (viewing, editing, downloading, etc.) is paramount.

This brief overview of some of the possibilities of handling data opened up by digitalising and Internet-enabling them indicates that there are ethical responsibilities around the way in which they are handled which go beyond ensuring anonymity, confidentiality and privacy. It is still a question towards whom (or what) these responsibilities are to be directed. This is not divorced from the question how the representational content of data is produced or constructed, as I go on to show in the next section.

Representational models

It is often the case that people who are in ethical decision-making positions – including designers and developers, researchers and their subjects, and the ethics committee members who must vet their research – assume certain models, or have certain implicit expectations of how representations work. Because the model or models they have in mind are not explicit, it is not clear to whom or what ethical consideration is due. To highlight the fact that data can be representations as well as simply bare information, while at the same time claiming – as I have – that this is an aspect that needs to be kept in the foreground while considering how or whether data should be digitally enabled and archived only goes a small way to broadening the scope of ethical negotiations around data. Representation is not representation *tout court*, as art history, literary theory and philosophical aesthetics attest. In the case of social science data, there could be competing assumptions and preconceptions about how representations function among different stakeholders. Different people will have a different implicit model about how representations operate, and this will be part and parcel of their ethical attitude towards data considered as a form of representation.

In this section, I describe five representational models, which are broad frameworks for the conceptualisation of data as representation, focusing in particular on the relation between representer/researchers and represented/data subjects. The models

are offered as heuristic tools for thinking about the link between representational and ethical considerations. It is not, however, the case that data sets will fall uniquely into one or other model; rather it is more usual that there will be a difference between the model that different stakeholders implicitly apply to the same dataset, and there may also be tensions and competition between them. The application of specific representational models can be affected by a number of factors, not least of which are the power relations between those in the representational context.

The representational relation defined by who represents whom is fundamental in determining representational content; among other things it also tells us much about the context of the representation, the expectations with which it was made, and the purposes. For example, a family snapshot has a different representational content to a passport photograph even if it looks the same. The representer/represented relation can be aligned with relations of power and the playing out of values in representations is often a matter of the distribution of power across the research environment – it is, after all, for that very reason that human subjects are protected by ethical regulations. This does not mean that power relations are always in favour of researchers; for example particular social groupings of potential data subjects can exert a great deal of influence in having some research but not others pursued (for example, research into so-called 'lifestyle' illnesses). In addition, researchers do not do research in a vacuum but are dependent upon research communities and institutions of different types.

The five models I consider are:

- Naturalism
- Isomorphism
- Figuralism
- Constructionism
- Interactionism

Naturalism

Naturalistic representation could also be labeled 'the disappearance theory of representation' as it prevails when representations do not seem to be representations, or

are so closely connected to what they represent that they seem as faithful as a representation can be. In these cases, the representations seems to be more like a screen or window and the surface of the representation seems to be a neutral 'giving over' to the thing represented, something that we simply look through in order to see the 'real' subject of representation. Naturalistic representation draws on resemblance theories of the way in which images and other visual media gain their representational content, and / or by the naturalistic variant thereof, which sees continuity between the way in which objects are perceived in the actual world and in art and other visualisations (eg a realistic picture of a face) (see for example Hyman 1997; Peacocke 1987).

In the case of photography and film, there is the further dimension of the causal connection that there exists between the colour, lighting and structural features (such as spatial arrangement) and the real world state of affairs of which they are the photograph or film. This creates the effect of a kind of immediacy in photograph and film: a capturing of something as it was at a particular time: the content of the photograph or film seems to be 'this-here-now'. The 'this-ness' and the '(t)here-ness' seem to be inscribed into these media in our experience of them (Barthes, 1981). This can also occur in recordings of voice, which can have an effect of presence in representation.

On a naturalistic view of representation, features of the subject are mapped onto features of the data. A face is recognisable in a naturalistic representation, particularly if it is a photograph or film. The data are also taken as reflecting aspects of subjects' persona: for example, their mode of dress, their gestures or other aspects of behaviour, and thus also to be revealing about 'thick identity'. It is part of the naturalistic model of representation that representational content is determined by the thing or object represented rather than by the representer, who – together with the technologies used to create and to present the representation – recedes into the background, or indeed does not figure at all.

Naturalism has many limitations as a theory of representation, since representations are not as neutral or transparent as naturalism suggests, and the representer as well as technologies of representation are an essential aspect of the way in which things get represented; *however it is the implicit 'theory' of representation which*

is most often spontaneously adopted by research subjects. Because of this, it cannot be discounted on the basis of a rival theory of representation. Indeed, because it is so deeply embedded in the phenomenology of representations, in which it is ordinarily not salient that representations are constructed in various ways, the concerns about the way in which subjects are represented should be taken very seriously indeed. On this model, because the representer's role is not phenomenologically salient, it can be held – in particular by subjects whose features can be immediately recognised in the data on the basis of resemblance – that the represented subject has the primary stake in the representational content, and ethical concerns are directed towards the subject of representation, with respect to both forms of identity, thin and thick. This is particularly the case because any 'representation as' content will appear naturalistically, and so as simply the way the subject is. If the sensitivities of a stakeholder who holds an implicit or explicit naturalistic model of representation are respected and acted upon, it will be important that discrete data should not be isolated from subjects, and the relationship between subject and data needs to be kept live. This would preclude re-use of data by researchers who have had no direct contact with the data subjects; it also precludes images or other representations being placed in data bases not agreed to by the subjects, as this will affect the way in which they are represented (even if they cannot be re-identified).

Isomorphism

This is an interesting view of the representational relation between patients and their digital medical records put forward by Eike-Henner Kluge (2001), and inspired by Wittgenstein's picture theory of meaning.

Kluge claims that since (1) digital patient records bring together items that in paper-based records would have remained separate; (2) they allow for a spatio-temporally 'joined-up' (in space and time) 'view' of the person qua patient, (3) combinations and manipulations are holistic (that is, a manipulation on one part affects the whole), digital health records are not simply copies of the features of the patient's health status, but actually share some features – such as the combinatorial features of representational elements) with the elements represented. This makes electronic patient analogues of patients. The analogical relation is limited to 'those aspects of the patient which have

generated data in the health care professional / patient encounter (Kluge, 2001, p. 30) and does not include the patient as a person.

This goes beyond naturalism as a theory of representation, because naturalism depends on a resemblance relation between discrete items (the subject's face and her image, for example). The analogical relation is instead a many-to-many relation, as it ranges over the diverse items in a digital patient record (blood tests, X-rays, diagnoses, etc.) and the diverse aspects of the patient to which they are connected. In particular, the analogical relation holds between the relations among the diverse items in the digital health record as a whole, and the relations among the diverse aspects of the patient. The analogical relation is thus a relation between two sets of relation. There is an analogical relation between these two sets of relation because they are isomorphic: that is, not just similar but the same with respect to their form or structure. A picture of a rectangle is isomorphic with a rectangular box because it has the same relation between (some of) its parts as the box.

According to Kluge, the ethical implications of isomorphism between digital health records and patients is that the first is an analogue of the second, and that they should be treated 'according to ethical principles that are analogues of the very same principles that normally should govern our conduct towards persons in the real world' (Kluge 2001, p. 39), and not merely heuristic codes of conduct. For example, it implies, according to Kluge, the ethical principle that 'The electronic patient record should be treated never as a mere thing but always as a person-analogue in information and decision-space.' (Kluge, 2001, p. 56).

Even stronger than naturalism, isomorphism sees data as being deserving of treatment analogous with that of the subject. We have already encountered the notion of reification in the context of some of the consequences of digitalisation. Reification is a feature of digital health records, according to Kluge:

> Once it has been generated, the existence and functioning of this patient analogue is independent of the patient and functions independently as basis for interventions and decisions-making. Therefore, to all intents and purposes, it has acquired what amounts to a functionally independent status. In this sense, we are

beginning to see the ontological reification of the patient record (Kluge, 2001, p. 33).

Thus, the digital patient record is itself a direct object of ethical concern, as is shown for instance by Kluge's Principle of Security:

> Data that have been legitimately collected about a person should be protected by all reasonable and appropriate measures against loss, degradation, unauthorized destruction, access, use, manipulation, modification or communication (IMIA Code of Ethics for Health Information Professionals).

To treat a set of data relating to a person as analogous to subjects is to treat it *as though* it is the subject. The data set contained in a digital patient record is the direct object of ethical concern, and the subject is the indirect object of ethical concern. This in no way implies that the subject is secondary to the data[6], but rather that the digital patient record mediates the ethical concern towards the patient.

This view of digital patient records has met with criticism. For example, one reviewer has criticised this account of digital patient records:

> Such a view of representation – apparently inherent to what Kluge refers to as 'an information-theoretic standpoint' – denies all the (error prone) choices, dilemma's, work, translations and (human) agency involved in the construction of records, and precludes appreciation of the well-documented need for a more empirically informed socio-technical understanding of the role, nature, and function of medical records in actual health care delivery (Van der Ploeg, 2003, p.66).

Isomorphism, thus, will be in tension with constructivism, in particular with ontological constructivism according to which that which is represented (the person, thing, event, etc.) is also a construct of research. This is just one of the tensions that may exist between competing representational models (a point to which I return).

---

[6] A further analogy, this time with theory of perception may help. Indirect realists do not hold that the real object indirectly perceived is secondary to the mental object directly perceived.

Isomorphism is a useful representational model because it allows a grasp of a particular feature of the Internet or Internet-related technologies, such as the Grid. Discrete items of data may not constitute items for either positive or negative ethical concern, but when several discrete items are connected up, they allow a picture of subjects to emerge which may well be an object of ethical concern, and will certainly have rich 'representation as' content (Nissenbaum, 1988; Nissenbaum 2004). The recent release by AOL of users search logs is an example: a single search is not ethically significant, but a series of searches is.[7] In the AOL case, Internet users were identified in some cases, but even in the absence of positive identification of particular subjects, on the isomorphic representational model there is reason to hold that the logs constitute an analogue of an actual person, and are therefore, as a body of data, a direct object of ethical concern, even if we do not know who the person is. We may owe these data, with the picture of a person that they present, the same kind of ethical concern that we owe to strangers, that is, to at least not harm, and in some cases try to protect.

Figuralism

A familiar topic in cultural studies, the rhetorical aspects of representation can be an unacknowledged force in the formation of representational content in the social sciences. However, the anthropologist Marilyn Strathern alludes to different figural constructions at play in the representational relation between the 'whole person' and an artefact:

> On the one hand, all over the world we find systems […] in which numerous distinct and specific items are drawn together, with the sense that nothing less than total enumeration will do. On the other hand, people can use the specifics as such to summon a larger vision of themselves. This idea is a different kind of access to totality: One artefact could be enough to point to the whole. […] Here one found another bifurcation, between wholeness imagined as the sum of parts and wholeness summoned in an individual item. (Strathern, 2004, p. 7)

---

[7] See for example M. Barbaro and T. Zeller, A face is exposed for AOL user 4417749, New York Times, Aug 9 2006,
http://www.nytimes.com/2006/08/09/technology/09aol.html?_r=1&adxnnl=1&oref=slogin&adxnnlx=1155
744807-DKt5sZmwY2uRl/mAsmGK9w

These are the rhetorical figures of synecdoche (a figure by which a more comprehensive term is used for a less comprehensive or *vice versa*; as whole for part or part for whole, genus for species or species for genus, etc. *OED*), for example, when the group or class to which a person belongs is used to denote the person, or the other way around; or metonymy (figure produced by substituting for a word or phrase denoting an object, action, institution, etc., a word or phrase denoting a property or something associated with it; an instance of this. *OED*), for example, when a feature associated with some item or person or group (religiosity, liberalism, money) stands for the item or person or group. Anthropological artefacts often function figurally in order to evoke or express aspects of the people from whom they emanate. A chalice is an anthropological artefact which expresses a great deal about Christian communities via the figure of synecdoche. Any number of other figural relations could play a part in determining the representational content of data; for example, a brain scan can be seen by a patient as a metaphor for her personality, and indeed artists, such as Susan Adlworth, have constructed self-representations out of them. They are significant relations in that they underlie the reasons why there is often an affective connection with data, and they are often powerful just because they are not obvious. They are sometimes a reaction or counter-reaction to reification, and they are part of the ways in which subjects or even researchers may include data in their 'stories' or their overall self-representation.

Figural representational content occurs alongside or as an overdetermination of representational content of data used for scientific purposes. Figures are a connotative rather than a denotative representational device, and they are an integral part of any symbolic system, linguistic or computational. Internet presentation has its own particular rhetoric, and figural content can be affected by the existence and placement of hyperlinks as well as by the possibilities of interaction that it allows for (Burbules, 1998). Figural representational content is often an essential part of 'representation as', and is not to be discounted on the grounds that it is subjective. Apart from the fact that this often rests on a false distinction between subjective and objective, relations of trust between researchers and subjects (either as individuals or as groups – as is borne out by anthropological gathering and use of data) may depend on understanding what data means for subjects, as well as what it could mean for others. In fact, this is an important methodological

consideration in the interpretation and presentation of some kinds of data, in particular qualitative data. Having said this, the ethical concerns that figural aspects of representations evoke can be very difficult to pin down, but in the context of so-called 'digital repatriation' and digital anthropology generally, politico-ethical consideration has in several cases left it to be decided by the provenance or origin of the data. In this case figural representational content – that is, metonymic or synecdochic figures as described above – may be an indication that the provenance of the data – the group or person from whom it was acquired – carries much and possibly over-riding ethical weight (Srinivasan, 2006; Carlson & Anderson, 2007). This is not only because of questions of ownership – though this may be the case – but also because it is the thick identity of the source of the data, that is at stake, and this identity is figurally indicated in the representation.

Constructionism

Data as construct is the oppositional term to data as naturalistic / isomorphic representation. In both of these, it appears that a representation is as it is because the subject it represents is as he/she is. A photograph shows a person with blue eyes because the person of whom it is a photograph has blue eyes. A constructionist representational model instead emphasises the extent to which representations are not neutral vehicles for grasping and conveying the features of the things they represent, but artificially assemble and produce representations. A huge research topic in its own right, constructionism in art and literature is nothing new and has even become the mainstream theory of representation.

In Internet research (of people on the Internet) constructionism is motivated in part by the diverse kinds of identities that are played out in the Internet, where there is a great deal of scope for people not being what they appear, and thus for any kind of naturalism or isomorphism to be extremely problematic (White, 2002, p. 249. And explorations of identity (in my sense of 'thick' identity) endorse the idea that subjects' identities are never simply clear-cut pre-existing entities, but are generated in an ongoing bricolage or construction. This is particularly apt on the Internet.

In Internet research, taking data to be constructs results in ethical concern shifting from the subjects who generated the representations to the representations themselves, for

example, showing them to be a certain kind of political, institutional, ethical entity by doing a discursive or political critique which shows how they have been constructed. This kind of reading can be critical of the subjects themselves or of the groups of which they are representative (White, 2002).

In e-research the constructs whereby subjects are represented are not generated by the subjects themselves, and thus raise a whole series of other issues. It is simplistic to say that they are generated by the researchers individually; rather they are a matter of the whole research context. And in e-research it is essential to take into consideration the role of the technology in constructing the data. Statistical research is a very good example of highly constructed research, since it takes data, processes it, analyses it into different aspects or measurements, and re-assembles or re-synthesises these measurements. There is nothing to say that constructs cannot be highly informative. E-research is likely to encourage constructed data, even from the very basic level of allowing for greater resources for annotating and analysing qualitative data. The result of these techniques are data which have naturalistic elements, but which as a whole, together with the annotations, codings, etc, are constructed data. The manipulation, standardisation, collection, and arrangement of data that digital and Internet technologies make possible will certainly reinforce the conception of data as construct – i.e. as not deriving its content from its relation to the subjects but from its relation to researchers and their context.

In e-research this model results in a shifting of the primary ethical concern from the research subjects to the data, conceived as constructs, and the primary stakeholders in that data become the researchers rather than the subjects. The acknowledgement of the data subjects consists in seeing them as the extrinsic occasion for the data rather than as intrinsically related to them, so 'representation as' does not actually reflect on the subject (even though it may seem to, in virtue of naturalistic content).

This is the attitude that allows for much actual practice in current social science and other research, where images and other information are regularly used by researchers in publications and presentations, and treated as their own property. The constructionist model can (and often does) live side by side with other models. For example, even

though on this model, data are constructed, there may also be aspects of naturalistic content or a perception that there is by data subjects. Researchers may have a sense of responsibility towards data and data subjects in virtue of naturalistic content, but these are seen as a matter of their own personal ethics (or 'Good Samaritan' ethics (Ess, 2002)) and not something which should be subject to external ethical protocols.

Unlike the isomorphic model which implies that data should be treated as an analogue of the subject, the constructionist model has no reason for doing so as this intrinsic link with the subject is broken. Rather the data are a constructed object of the researcher (and their research context), and constructions are not the same kinds of entities as analogues. Constructions imply relations of ownership, material and intellectual, rather than relations of identity, and indeed seeing the representational content of data as constructed by the researcher brings this closer in line with legal interpretations of the ownership of data and copyright, wherein the processes of creating data and data collections are taken into consideration in deciding ownership (D'Agostino et al., 2006; Burk, 2007). Thus the sensitivities around data informed by this model would indicate that the data themselves are a direct object of ethical concern, but insofar as they are an object of ownership by the researcher; researcher in turn mediates ethical concerns towards the research subject.

Thus on this construal, obstacles to sharing data for collaborative research or re-use will have more to do with the research interests of the researcher, rather than with data subjects. Constructionism is a view of data as representation which takes onto itself the burdens of ethical qualms associated with data sharing, and reification, re-contextualisation, manipulation and the overall way in which subjects' identity is represented, and much is a matter of researchers' ethical sensitivities. This is the underlying model held by emphasis placed on researchers' negotiating archiving of data on behalf of data subjects.

Interactionalism

On an interactional account of the representational content of data it is a mistake to see data as representing the subject(s), or as being a representation of the subject(s) by researchers. Data are generated in the interactions between subjects and researchers,

within particular contexts and settings. In addition, on this view data are not only *of* someone, but always addressed *to* someone. The relation to the addressee specifies much about the data. For example Heath & Luff (2000) show that the features of medical records (what is recorded, what is left out, how it is recorded) are motivated by the fact that they are addressed to other professionals. This is in line with dialogical theories of language, such as that of Bakhtin and followers, which highlight the fact that the way in which utterances are directed towards their addressees shape their form and meaning. With this conception of data, the technologies involved in generating the data could be very important in defining the interactions around it, because technologies have specific sets of interactions built into them. Thus the possibility of collaborative interpretation of data on the grid (for example using MiMeG) will make a difference to how it is gathered, and to interactions with the data subjects, and this possibility is thus always a part thereof.

The representational content of data cannot be understood at the level of individual subjects or researchers, but rather as arising through ongoing interactions among members of a group within a context. The context ranges over immediate concrete context (the physical features of the interview situation for example), to broader contexts (social and institutional). Importantly, the content of data is seen as emergent in that it arises in the interaction (Goguen, 1997). Another important feature of this model of representation is that it incorporates the values which people hold as an integral aspect of the data that emerges out of the research interaction. For example, in ethnomethodological research which is one expression of the interactionist model, it is stressed that communication and information have an 'inalienable ethical dimension' (Goguen, 1997, p. 47), and are not abstract ideas or ideals existing in a realm separate from data and research. This means that values are embedded within the representational content of data. On this account, whether data could be shared by being collaboratively analysed would depend on the ethos of the interaction. If trust is central to the interaction, the data may be shared with others who share in the trust relation (perhaps other researchers who have the data collectors trust). However, this conception of data also means that it would be methodologically unsound to de-contextualise data from the

interactions and contexts in which it was generated. Thus the possibility of re-use in contexts entirely dislocated from the primary context is not great.

On this model, the data are the direct object of ethical concern on the part of both the researcher and the subject. The researcher cannot arrogate to him-or herself the role of mediating the ethical concerns of the subject. The subject as well as the researcher has responsibilities towards the data. There are different degrees to which this model can be put into practice. It is often used together with constructionism, but it also lends itself to a participative model of research. For example, the 'Enabling Diversity: Extending Collections Information with Arctic Communities' project under the auspices of the Cambridge University Museum of Anthropology and Archaeology is using the greater interactive capacities of the Internet, and in particular online social computing (Web 2.0) to allow user communities to interact directly with the collections, and specifically 'to add information and develop historical narratives (Corti, 2000).

The interactional model allows data subjects to be active participants in their self-representations, either as individuals or as members of a community. This exploits the full potential of digital and Internet technologies to act as facilitators of reflection on these representations and the way in which they function in the formation of identity.

This is no doubt not a complete list of possible representational models for e-research data. Missing from this picture of representational relations is the role of technologies involved, and their designers and developers, and undoubtedly ongoing study will show these to have a very significant role. However, it should be clear from the discussion that the primary objects of ethical concern shift according to the implicit representational model held by different role players: researchers, subjects and archivists.

The models may co-exist, but not always peaceably: for example, there is likely to be tension between naturalism and isomorphism on the one hand and constructionism and interactionism on the other. Differences between implicit representational models held by different role players could account for some of the struggles around data and the different attitudes towards the different tools and technologies for e-research in the social science domain. A predominantly constructivist model of data may find data archiving and data mining less problematic than a naturalistic or interactional model. For example,

data which have a predominantly interactional representational form is less likely to be apt for data archiving, and this particular technology may be resisted by researchers and participants alike. An option would be to exploit interactive Internet technologies in order to extend this to archived data.

These representational models have been outlined in a more-or-less neutral way, in that no attempt has been made to weigh one up against the other, epistemologically or methodologically. The models outlined (and there may be others) are often implicitly held within domains where decisions and choices are made concerning ethical aspects of data, and need to be 'teased out' of the domain in order to understand how these choices and decisions are being framed by all participants, be they researchers, users, developers, funders, or ethics committee members. A further role for attempting to understand which representational models are at play is for necessary conversations to be had about ethical aspects of Internet-enabled research involving data about human subjects. The models are thus but heuristic tools to probe attitudes towards representation, bring them to the surface, and to tackle the ethical aspects of data as a form of representation. It is important that the conversation not end at the point of ensuring anonymity and confidentiality.

Conclusion

This article has presented a case for considering data as representation and not only as information as we attempt to grapple with the ethical implications of e-research. A distinction was drawn between thin and thick identity, and it was suggested that thin identity – the possibility of re-identifying an individual – leads to ethical considerations around anonymity, confidentiality and privacy. These are traditionally the ethical issues seen to issue from a conception of data as information. Thick identity instead has to do with the experienced personhood of data subjects, and is a question of the way in which subjects are represented by data. These representations are a matter of ethical concern as they are the arena in which implicit personal and social values which inform moral judgments, are expressed and played out.

The paper then outlined what possibilities of handling data are made available by the digitisation, large-scale data-bases and other computational resources associated with

e-research. These include standardisation, reification, manipulation, and the analytical and interpretive capabilities of annotation, tagging, data-mining and visualisation.

In the face of the technological possibilities for handling digital data presented by e-research, researchers using data about human subjects have a further set of ethical issues to consider with regards to the representational content of data. These issues are complicated by the fact that it is not simply what – or who – is 'immediately' represented that needs to be considered, but also the implicit representational model or theory which goes some way to explaining why different people involved with data may feel that they have a particular stake in data. The third part of the paper outlined five different representational models: the naturalistic, isomorphic, figural, constructionist and isomorphic model. While not an exhaustive or definitive list, the models are put forward as a way of highlighting competing concerns and interests around data that can be traced back to the role of different roleplayers in determining representational content, as these may be perceived either by subjects or by researchers following different methodologies and research strategies.

The understanding of data as representation presents a challenge to the designers and users of e-research tools and technologies that go beyond the control of access and ensuring of confidentiality and anonymity. It is not possible to apply a cookie cutter model of ethical conduct with respect to Internet-enabled data relating to human research participants. Rather, it is important to try to understand what models of representation are at play for researchers and research participants, what negotiations are required around these models, and how technologies can be shaped to build in awareness of the ethical aspects of representations of people in Internet-enabled data.

References

Bamberg, M. (2004). Narrative discourse and identities. In J.C. Meister, T. Kindt, W. Schernus & M. Stein (Eds.), *Narratology beyond literary criticism* (pp. 213-238). Berlin and New York: Walter de Gruyter.

Barbaro, M. & Zeller, T. (2006). A face is exposed for AOL user 4417749. *New York Times*. 9th August, 2006. Retrieved January 15, 2008 from http://www.nytimes.com/2006/08/09/technology/09aol.html?_r=1&a dxnnl=1&oref=slogin&adxnnlx=1155744807DKt5sZmwY2uRl/mAsmGK9w.

Barthes, R. (1981). Camera lucida: Reflections on photography. New York: Hill & Wang.

Brey, P. (2000). Disclosive computer ethics. *Computers and Society, 30*(4), 10-16.

Burbules, N.C. (1998). Rhetorics of the Web: Hyperreading and critical literacy. In I. Snyder (ed.), *Page to screen: Taking literacy into the electronic age* (pp. 102-122). London and New York: Routledge.

Burk, D. (2007) Intellectual property in the context of e-Science. *Journal of Computer Mediated Communication*, *12*(2), article 13. Retrieved December 20, 2006, from http://jcmc.indiana.edu/vol12/issue2/burk.html.

Carlson, S., & Anderson, B. (2007). What *are* data? The many kinds of data and their implications for data re-use. *Journal of Computer-Mediated Communication*, *12*(2), article 15. Retrieved January 15, 2008 from http://jcmc.indiana.edu/vol12/issue2/carlson.html.

Corti, Louise (2000). Progress and Problems of Preserving and Providing Access to Qualitative Data for Social Research—The International Picture of an Emerging Culture [58 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [Online Journal], *1*(3). Retrieved Janurary 15, 2008 from http://www.qualitative-research.net/fqs-texte/3-00/3-00corti-e.htm.

D'Agostino, G., Hinds, C., Jirotka, M., Meyer, C., Piper, T. & Vaver, D. (2006). On the importance of intellectual property rights for eScience and integrated health record. *Integrated Healthcare Workshop*. Edinburgh, UK.

Ess, C. (2002). Introduction. *Ethics and Information Technology 4*(3), 177-188.

Ess, C. & AoIR ethics working committee. Ethical decision-making and Internet research: Recommendations from the aoir ethics working committee. Approved by AoIR, November 27, 2002. Accessed January 15, 2008 from http://www.aoir.org/reports/ethics.pdf.

Geertz, C. (1976). Art as a cultural system. *MLN 91*(6), 1473-1499.

Goguen, J. (1997) Toward a social, ethical theory of information. In Bowker, G.C., Turner, W., Gasser, L. (Eds.), *Social science, technical systems and co-operative work: Beyond the great divide* (pp. 27-56). Mahweh, NJ: Lawrence Erlbaum.

Gombrich, E.H. (1983). *Art and illusion*. Oxford: Phaidon.

Goodman, N. (1976). *Languages of art*. Indianapolis: Hackett.

Gross, L., Katz, J., & Ruby, J. (2003). *Image ethics in the digital age*. Minneapolis & London: University of Minnesota Press.

Habermas, J. (1988). *Le discours philosophique de la modernité*. Paris: Gallimard.

Heath, C. & Luff, P. (2000). *Technology in action*. Cambridge, England: Cambridge University Press.

Hinds, C., et al. (2005). Ownership of intellectual property rights in medical data in collaborative computing environments: *First International Conference on eSocial Science*. Manchester, UK.

Hyman, J. (1997). Words and Pictures. In Preston, J. (Ed.), *Thought and Language* (p. 42). Royal Institute of Philosophy Supplement: CUP.

Jirotka, M., et al. (2005). Collaboration and trust in healthcare innovation: The eDiaMoND case study. *Computer Supported Cooperative Work, 14*, 369-398.

Kindon, S. (2003). Participatory video in geographic research: A feminist practice of looking? *Area, 35*(2), 142-153.

Kluge, E.W. (2001). *The ethics of electronic patient records*. New York: Peter Lang.

Nissenbaum, H. (1988). Protecting privacy in an information age: The problem of privacy in public. *Law and Philosophy, 17*, 559-596.

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review, 79*, 119-157.

Peacocke, C.A.B. (1987). Depiction. *Philosophical Review, 96*, 383-410.

Srinivasan, R. (2006) Where information society and community voice intersect. *The Information Society*, *22*, 355-365.

Strathern, M. (1999). *Property, substance and effect: Anthropological Essays on Persons and Things*. London & New Brunswick: Athlone Press.

Strathern, M. (2004). The whole person and its artifacts. *Annual Review of Anthropology, 33*, 1-19.

Tavani, H. (2004). Internet privacy: Some distinctions between Internet-specific and Internet-enhanced privacy concerns. *Ethicomp Journal 1*(2). Accessed January 15, 2008 from http://www.ccsr.cse.dmu.ac.uk/journal/abstract/tavani_h_internet.html.

Tomaselli, K. G. (1996). *Appropriating images: The semiotics of visual representation*. Højbjerg: Intervention Press.

Van Der Ploeg, I. (2003). Review of Eike-Henner Kluge. *Ethics of Electronic Patient Records, 5*(1), 66-67.

White, M. (2002). Representations or people? *Ethics and Information Technology, 4*(3), 249-266.

Researching the Researchers:
Market Researchers, Child Subjects
and the Problem of "Informed" Consent

S.M. Grimes
Applied Communication and Technology Lab
3520-515 West Hastings Street
Simon Fraser University
Vancouver, BC
Canada V6B 5K3
smgrimes@sfu.edu

Abstract

Children's online activities raise a number of ethical challenges for academics, market researchers and other media professionals. Public debates over privacy and commercialization within popular children's websites have led to the establishment of a small number of governmental initiatives aimed at regulating online data-collection involving children. To date, however, preliminary responses have been ineffective in providing ethical standards of practice to the online children's marketing research industry, wherein commonly accepted rules for conducting research on children, including informed consent, seem to be bypassed altogether. The resulting disjuncture between these practices and academic ethical standards causes unexpected challenges for researchers, wherein rules intended to protect children could inadvertently exclude studies of the more questionable facets of market research.

Introduction

As digital technologies continue to blur traditional boundaries between public and

private, social conceptions of childhood as a separate and protected domain are eroding.

Whereas children have historically been excluded from directly participating in most

aspects of the public sphere—from voting to mass media production—the Internet now

offers users of all ages and ability a variety of opportunities for participation in the collaborative construction of a highly-fragmented, yet wide-reaching, digital public sphere. More than any other group, children and youth have flocked to these opportunities, often becoming society's earliest and most enthusiastic adopters of online communication tools (chat, MSN), social-networking sites (*Club Penguin*, *Bebo*), and collaborative cultural experiences (online games, mash-ups, file sharing). Yet, children's involvement in online culture has also raised a number of new challenges for families, governments, academics, and media professionals. Children are not always able to make responsible and informed decisions about Internet content and interactions, and they are not always equipped to navigate the increasingly ambiguous boundaries between private and public life online (Turow, 2001; Turow, 2003; Shade et al., 2004). Younger children in particular are vulnerable to exploitation, manipulation and harm as a result of unregulated exposure to the darker aspects of the World Wide Web. Children have thus become the locus of heated debates and tensions about the role of digital technologies in Western society—representing both the hope that the Internet will open up possibilities for democratic cultural participation, as well as the fears associated with lowering the domestic floodgates to the unrelenting access of the digital world.

In North America, children spend the vast majority of their time online on commercially owned and operated websites (Seiter, 2004; Moore, 2006). Despite the vastness of the Internet, the number of children's sites with no advertising or branding is almost negligible. As Neuborne (2001) reports, "[T]he number of children's sites with no advertising dropped from 10% of all kids' sites [in 1999] to just 2%" in 2000 (p. 108). As a result, one of the most prevalent new issues to emerge from children's increasing Internet use is the way in which it enables a deepening relationship between child users and commercial entities (Montgomery, 2000). Every day, children enter into complex relationships with the adults who create and manage profit-driven online spaces directed at kids. These include business relationships, consumer relationships, and legal relationships, most of which occur 'behind-the-screen' of seemingly entertainment-oriented spaces. For while many of these websites provide children with hours of online fun, vibrant communities of interest, as well as access to and information about beloved media characters (Aikat, 2005), they also operate as important marketing tools (Seiter,

2004; Grossman, 2005; Steeves & Kerr, 2005). Many of the most popular commercial kids' sites function primarily as forms of interactive advertising and branding, as well as venues for market research—wherein the unique levels of access enabled by the Internet are used to gather hitherto unimaginable amounts of personal information, thoughts and opinions from children (Kapur, 1999; Montgomery, 2000; Steeves, 2006).

A growing public debate around issues of privacy within commercial children's websites has slowly led to the establishment of a small number of governmental and institutional initiatives aimed at regulating online and electronic data-collection involving children. Most notable among these is the *Children's Online Privacy Protection Act* (COPPA) in the United States, and the *Personal Information Protection and Electronic Documents Act* (PIPEDA) in Canada, both of which establish nationally-enforced restrictions on the collection of minors' personal information (such as name, address, postal code). Yet, market research in children's websites and online communities persists, and a growing body of literature now suggests that inadequate industry regulation and loose ethical standards have led to the emergence of highly questionable standards of practice among at least some online market researchers (Kapur, 1999; Klein, 2000; Montgomery, 2000; Linn, 2004; Seiter, 2004; Rushkoff, 2006; Nairn, 2006). From surveys and polls disguised as "personality quizzes," to behind-the-screen surveillance of online activities and communications (Chung & Grimes 2005), many children's websites contain features that enable the compilation of vast databases of non-personally identifiable, and thus unprotected, information about child users. The resulting aggregated data is then sorted and made sense of (and sometimes sold) in the form of detailed "youth trend" reports (Grimes & Shade, 2005), which are highly valuable to those involved in the production and promotion of children's consumer goods. These highly invasive, yet conspicuously covert, practices raise new questions about the scope and limitations of children's privacy (both in terms of expectations, as well as the protection thereof), intellectual property, and media literacy, which have yet to be adequately addressed within academic and policy discourses.

Like so much privately-funded, proprietary research, only a fraction of the studies conducted by the children's industries is made available to the public (Kline, 1993; Schor & Ford, 2007). While large-scale market research reports, such as trend and demographic

reports, are occasionally made available for purchase—as published reports or as trade publications—companies that conduct market research internally have little incentive to share their methodology and results with other researchers. Academics attempting to study the children's market research industry thus often grapple with limited access to findings and a lack of information about the research methods used. Examples of recent approaches used by scholars include retroactive analysis of published reports and marketing literature (Kline, 1993; Cook, 2000), participation in marketing industry conferences (Quart, 2003; Linn, 2004), and interviews with marketing professionals willing and able to share trade secrets (Klein, 2000; Schor, 2004; Rushkoff, 2006). With the integration of online techniques, however, new possibilities for more direct forms of observation have emerged, allowing academics to track and witness certain market research activities firsthand.

This paper provides an overview of the ethical challenges involved in conducting an academic study of some of the market research practices described above, and the difficulties encountered while attempting to secure ethical clearance to conduct the research. While a detailed review of the study itself is beyond the scope of this paper (as well as published in full elsewhere—see Grimes and Shade, 2005; Chung & Grimes, 2005; Grimes 2007), the first section will establish the contextual background of this discussion by providing a brief description of the project, its key findings, and the themes that emerged during analysis. The subsequent section will review the regulatory frameworks within which contemporary online market research practices operate, including relevant government policies and industry guidelines. I will then discuss how this framework differs from the ethical standards required of research conducted within the context of a Canadian university, highlighting points of contention that could problematize, and potentially disrupt, academic inquiry into market research practices (particularly those involving minors). The third section consists of a case study illustrating the disjunction that exists between the ethical standards applied by industry and those applied within a specific academic institution, and the impact this can have on the ethical clearance process and research design. The final section will discuss unresolved issues and ethical questions that arose during the case study, and highlight

persisting gaps within current academic standards of practice for researching children online.

Terms of Service, Terms of Play

Children's newfound role as user-producers of Internet content is often assumed within the context of commercial websites ("Kids Ages 2-11," 2004; Steeves, 2006; Prescott, 2007; Richtel & Stone, 2007), which implicate them in a complex network of business and legal relationships. In order to determine the nature and scope of these relationships, an in-depth content analysis was conducted of game-themed.[1] websites ranked among the "top ten" most frequented by children in 2002 and 2003 (Grimes, 2007). Sixteen case studies[2] were selected based on web-traffic statistics released by industry research firms Hitwise (Greenspan, 2003) and Nielson/Netratings ("Nearly 20 Percent," 2002). The content analysis included a comprehensive survey of the activities, features and information available on each site, which included preliminary discourse and textual analyses of the sites' contents. Where applicable, a step-by-step record was made of any member registration processes required to participate in the sites, and of additional features or activities that these processes enabled. The presence of interactive and user-generated features was recorded, and a random sample was observed. This included reading a small number of forum posts and user-created webpages (where available), participating in surveys and polls, and playing a representative sample of games on each site. The content analysis also included observation of the presence, positioning and contents of relevant texts such as privacy policies and TOS contracts. The TOS contracts were then analyzed in depth, using Russo's (2001) checklist of "15 significant points" to look for when reading TOS or EULA contracts.

---

[1] Game sites were specifically chosen to reflect the activity that children report engaging in and enjoying the most while online. Industry analyst NPD Funworld recently announced that 45% of "heavy gamers" and nearly one-third of "avid console gamers" are between the youthful ages of 6 and 17 (Graft, 2006). In 2003, 87% of children aged 7 to 12 years reported "playing online games" as their favourite online activity (Greenspan, 2003), and all five of the "top five" online destinations most visited by children aged 2 to 11 featured online games. More recently, Roberts, Foehr and Rideout (2005) found that children aged 8 to 18 years spent more time playing online games than on any other online activity (including email, instant messaging and chatrooms).

[2] Yahoo!Games: Yahoo! Fantasy Sports Baseball; Yahoo! Fantasy Sports; Yahoo! Fantasy Sports Football; EA Online; Pogo (EA); MSN Game Zone; Kraft Entertainment; Neopets; EverythingGirl: Diva Starz; EverythingGirl: Polly Pocket; EverythingGirl: Barbie; Disney Online; Disney's Toontown Online; gURL.com.

Among the major trends identified[3] was a predominance of "advergames"—games containing themes, activities or images that directly related to a specific product(s) or brand(s)—some of which consisted of little more than brand-recognition exercises. A second key trend among the sites reviewed was the collection of personal information, such as email address, date of birth and gender, although some also asked for names and home address identifiers (such as state/province, country, or zip code/postal code). Although the current industry enthusiasm for "Web 2.0" applications had not yet occurred at the time of the study, another key trend identified was an emphasis on "user-generated content." The vast majority included some form of social software or community-building tools (for example, multiplayer components, forums or chat rooms, or even email and e-card services), and more than half allowed participants to submit content (such as game reviews, poetry contest submissions or fan art). Two-thirds of the sites solicited players to complete polls or surveys, or to customize some aspect of the site or gameplay. These same sites often featured surveys, polls or customizable features directly relating to (and soliciting customer opinion about) particular products or brands.

The patterns uncovered during this study are consistent with previous work in this area, which has found market research taking an increasingly central role in both the creation and manipulation of children's digital content (Montgomery, 2000; Seiter, 2004; Steeves & Kerr, 2005; Moore, 2006). For example, the prominence of polls and other overt forms of market research have been noted by Seiter (2004), who describes how children's "interests, habits, and abilities in the online environment [have become] the subject of intense interest by marketers" (p.93). Similarly, Montgomery's (2000) overview of kids' online culture describes how the "intense focus on research within the new media industries has produced a wealth of information, much of it proprietary, which is guiding the development of digital content and services for children" (p.638). In their study of children's online play and chat activities within commercial websites, Steeves and Kerr (2005) found marketers "artificially manipulating the child's social environment and communications in order to facilitate a business agenda" (p.91). For example, the *ELLEgirl.com* site used an automated system to guide chat discussions towards specific

---

[3] For a full description of the features and contents of the websites examined, please see Grimes, 2007. For a detailed account of the methodology and content analysis protocol, see Grimes, 2005.

products, or to ask users about consumer preferences (Steeves & Kerr, 2005). The utility of "child-generated content" for market research purposes has also been explored in depth within the marketing literature (Lindström, 2003; Solomon & Peters, 2005; Bennett, 2006; Helyar, 2007), which often legitimate covert research strategies by emphasizing the opportunities for consumer "empowerment" and cultural participation that these applications may concurrently provide.

For the most part, the market research strategies utilized within children's commercial websites are rarely divulged in full to their users. While some of the sites included in the analysis identified a small number of features as "advertising content," little to no explanation was provided that certain features—such as polls and surveys, for example—could be used for advertising and market research purposes. Instead, these were presented as entertainment (such as "personality quizzes"), as a way to earn the "currency" or points needed to participate in the sites, or as a way for users to "help out" (e.g. sharing information to keep the site free, or as a way to improve the site's content). As for more covert practices, like data-mining and chat room ethnography, children are rarely (if ever) warned that their contributions could provide valuable information to market researchers. Yet, it is wrong to presume that children's digital literacy extends to these types of activities. Studies of young children reveal significant limitations in their overall ability to think critically about online information and activities (Kline, 2001; Livingstone, 2006). This in turn is supported by longstanding research showing that younger children are not always able to differentiate advertising from programming content in traditional media forms such as television (Preston, 2004). Child development experts generally agree that children under eight years of age are "prone to accept advertiser messages as truthful, accurate and unbiased" ("Television Advertising," 2004). When faced with a choice to respond to an online survey that promises to "help us make the site more fun for you!," for example, younger users may not have the critical skills to recognize the survey as market research, or to understand that their answers could be used to target advertisements at them later on.

Furthermore, emerging research shows that younger children often do not understand common "e-business" practices, such as the use of "Cookies" to track users online (Shade et al., 2004). They also do not always comprehend the full repercussions of

divulging personal information online, which may in part explain why children are more likely than adults to give out sensitive information, particularly in exchange for a free gift or reward (Turow, 2001). On the other hand, Livingstone (2006) argues that children deeply value privacy, and often use the Internet to "carve out private spaces" and experiences as respite from their otherwise "increasingly restrained lives." As Livingstone describes, "Our findings on what children say and do online suggest a series of 'uses of privacy'. These include feeling in control, mastery of the situation, something they may experience less when under the surveillance of an adult gaze" (p.137). While the "adult gaze" of the marketer is substantially different from that of a parent or teacher, it is unclear that children would feel as comfortable contributing to spaces they knew were under continuous adult monitoring and analysis.

The majority of children's websites seek to establish the parameters of their relationships with users through website privacy policies and TOS contracts. A user's consent to abide by the rules established by the privacy policy and TOS contract is assumed once he/she begins using the site. In the vast majority of the TOS contracts reviewed, a stipulation was included that any and all user submissions would automatically become the exclusive and unlimited property of the site (the only exceptions were the *Yahoo!Games* sites, which were also some of the only sites that did not provide space for user-generated content). In most of the cases reviewed, the types of user submissions and breadth of copyright assumed were both broadly defined—even though the user was rarely acknowledged as the original owner of her/his submissions in the first place. Significantly, it was not standard practice to include detailed information about marketing research practices in either the TOS contracts or the privacy policies unless these directly involved "personally identifiable information" (such as name, address, Social Security number).

Within the realm of adult users, TOS contracts and end-user license agreements (EULAs) are at the centre of ongoing debate and legal conflict (Grimes, 2006). These debates have challenged the validity of the sweeping terms and wide-ranging intellectual

property (IP) claims often included in these documents.[4] Most recently, a similar conflict has surfaced within the context of Massively Multiplayer Online Games (MMOGs). While the EULAs of these games often require that players consent to transfer and waive all subsequent claims to copyright over their in-game creations and communications, player communities, legal experts and scholars are challenging the legality of these terms and advocating that players should retain certain rights of authorship over their contributions (Castronova, 2003; Dibbell, 2003; Lastowka & Hunter, 2004). For instance, Taylor (2002) argues that online environments, such as MMOGs, are also spaces in which individuals invest a significant amount of time congregating, creating avatars, producing cultures and communities, sharing in leisure activities, and reproducing economies. In terms of the characters and items players acquire and create through the process of gameplay, Taylor (2002) writes that players are at the very least the collaborative authors (and hence partial owners) of any cultural artefacts that result from their efforts:

> It takes a player to create a character and it takes the time of the player to develop the character. Through their labor they imbue it with qualities, status, accomplishments. Indeed, while the owners of a game provide the raw materials through which users can participate in a space, it is in large part only through the labor of the players that dynamic identities and characters are created, that culture and community come to grow. (p.232)

While the MMOG debates have not yet considered how these same arguments might apply to child players, there is no immediate reason to exclude children's in-game submissions from discussions of authorship and the significance of users' creative contributions. Like the adult players described by Taylor (2002), children also produce a substantial amount of content when the necessary features are made available to them. For example, *Neopets.com* contains over 12 million pages of content created by its members, 39% of whom are under 13 years of age ("Frequently Asked Questions," 1999-

---

[4] It should be noted that many MMOGs require the player to click "Agree" to the EULA upon installation. However, although some require the player to scroll down to the bottom of the document, the length of the contracts (which are often between 10 and 20 pages) and their heavy use of jargon can be seen as likely deterrents to players taking the time to read and fully understand the terms to which they are agreeing.

2006). Issues surrounding children and authorship in the digital context, and the intellectual property that might result, have not yet been addressed within academic and legal fields. However, the notion that child-generated content has some form of tangible value is readily apparent within the TOS contracts' sweeping copyright claims, as well as the revenue generated from market research reports.

Unlike the case of the MMOG debates, TOS contracts found on children's websites become vulnerable to a special type of challenge, in that contracts made with minors are voidable and unenforceable in North America.[5] This raises serious questions about the applicability and validity of the claims and terms made within TOS documents *vis-à-vis* child users. As with questions of authorship, however, the validity of the minors' contracts found in commercial websites has yet to receive public attention. In the meantime, children's websites appear to operate as though the contracts were legal and binding.[6] In the absence of governmental or official industry policies regulating children's non-personally identifiable information, the privacy policies and TOS contracts currently fulfill a sort quasi-regulatory function that has not only become common practice within the children's digital environment, but could eventually set the tone for future, formalized regulatory initiatives. As Coombe (1998) argues, "People's anticipations of law (however reasonable, ill informed, mythical, or even paranoid) may [eventually come to] shape law and the property rights it protects" (p. 9).

The content analysis of the gamesites' TOS contracts revealed a number of additional problems relating to research ethics and participant rights (Grimes, 2007). The TOS contracts reviewed were difficult to find, often only accessible by clicking a link at the bottom of the page. They often consisted of lengthy texts (ranging between 3 and 12 single-spaced printed pages) that made heavy use of legal terminology, jargon, and long,

---

[5] In Canada, contracts with children can be voided at their request if they are not beneficial to the child. The only exception, as established in Miller v. Smith & Co., [1925] 2 W.W.R.360, 377 (Can.), is in instances where the child enters into a contract to obtain the "necessities of life" (food, clothing, shelter, etc.). For US legislation on minors' contracts see California Family Code 6710 (West 2004); Sparks v. Sparks [1950] 101 CA2d 129 (Cal.); Burnand v. Irigoyen [1947] 186 P.2d 417 (Cal.); Scollan v. Gov't Employees Ins. [1963] 35 Cal. Rptr. 40, 41 (Cal. Dist. Ct. App.); and Mitchell v. Mitchell [1998] 963 S.W.2d 222, 223 (Ky. Ct. App.) (Hruby, 2006).
[6] Surprisingly, minors' rights and special legal status have thus far been excluded from the adult-oriented MMOG debates, which tend to focus on T-rated games that have traditionally been played primarily by adults (Yee, 2006).

convoluted sentences. Generally, only a minimum effort was made to make the contents of these contracts accessible to children—in some cases, players were not even directly instructed to read the TOS, assuming instead a prior knowledge and experience of contracts that many children do not have. Only two of the sites reviewed provided a child-friendly version of the TOS (Kraft/Postopia and Neopets), both of which abridged and over-simplified important clauses of the full-text document. It is highly doubtful that children can legitimately be expected to have the skills and knowledge required to understand the concepts and implications of many of the clauses included—assuming, of course, that they are inclined and able to read through these difficult texts in the first place.

It is here that the issue of "consent" becomes significant. Recent studies suggest that children rarely read privacy policies (Sandvig, 2000; Turow, 2001), which they find "long and boring" (Burkell et al., 2007). Furthermore, it is questionable that young children have the skills and knowledge required to understand the contents of policies and legal documents without adult assistance or child-friendly explanations. While little research has been conducted on children's understanding of website policies and contracts, studies of adult Internet users reveal that only a small proportion understand the legal implications of privacy policies (Turow, 2003). Studies of children's understanding of economic concepts show that while rights of use and control are relatively easy for children to grasp, other concepts, such as the right of transfer, are much more difficult for children to comprehend (Cram, Ng & Jhaveri, 1996). For example, it is not until the age of 11 years that notions of private ownership and the right of transfer are fully understood by the majority of children. For very young children (under 6 years of age), Furnham (1996) argues, "[M]ost economic events are still simply observed and accepted as mere ritual" (p. 31). With this in mind, it is easy to see how clicking "I Agree" to a TOS contract can become a ritualized part of children's online experience, rather than a true indication of consent.

In terms of parental consent, there appears to be no standard framework for ensuring that parents are aware that kids' online activities are used for research purposes (Turow, 2001), or that verifiable informed consent is granted. In the few instances where some form of parental consent was sought—for the child's participation in the site or in

research where personal information was gathered—no evidence was found during the content analysis of a consistent strategy for ensuring that an *actual* parent was the one granting the consent. The websites reviewed were thus found to be doing very little to a) ensure that child users were both properly informed and had given explicit consent to present *and* future participation in market research; or to b) ensure that the child's parents had given informed consent for their child's online interactions to be used for market research purposes. It remains questionable that children and parents are ever truly informed of the full extent of corporate and legal mechanisms at work within many commercial websites and online games. This highlights the growing need for an in-depth examination of the issues surrounding informed consent in children's commercial websites, and its relationship to the academic research context.

"Business as Usual" vs. Research Ethics

Marketing research in Canada is regulated by a combination of federal, provincial and industry initiatives. In terms of federal regulation, PIPEDA provides a number of restrictions around the collection, use, and disclosure of personal information within the private sector. One of these restrictions includes a warning that seeking consent for data-collection "may be impossible or inappropriate when the individual is a minor" (Clause 4.3 Principle 3). However, the policy fails to provide any guidelines or restrictions specific to the special needs and vulnerabilities of minors, or to acknowledge the vast variability in aptitude and legal responsibility that exists among children and youth. Provincial policies, such as Quebec's *Loi sur la protection de la consommateur*, also enforce restrictions on such activities as using information from draws and contests for marketing research. Since a large proportion of popular children's websites originate from the US, many sites must conform to the Federal Trade Commission's (FTC) *Children's Online Privacy Protection Act* (COPPA) (2000). As with PIPEDA, however, the COPPA regulations only apply to "personal information" (defined as any identifier that permits identification or physical contacting of a specific individual) and personally identifiable information (which includes any information gathered from the child when combined with a "personal information" identifier) (Section 1302, p. 8). Children's non-

personally identifiable content, which includes all other types of content, does not receive any special protection in either Canada or the U.S.

The industry's self-regulatory guidelines do not provide much additional protection for children's content. The Canadian Marketing Association's (CMA) recently updated *Code of Ethics and Standards of Practice*, for example, requires that "all marketing interactions directed to children that include the collection, transfer and requests for personal information require the express consent of the child's parent or guardian" (Section K3) for participants under the age of 13 years. Relevant US-based associations, including the Marketing Research Association and the Interactive Marketing Research Organization, merely require that members comply with COPPA requirements. As with the COPPA and PIPEDA requirements, industry guidelines only provide safeguards to personally-identifiable information, such as names, addresses, and telephone numbers, with little consideration for the various other types of information that market researchers often find most useful (Smith & Clurman 1997; Sutherland & Thompson 2001; Lindström 2003). For example, COPPA only protects consumer preference data, such as hobbies, interests, and Cookie-enabled behavioural tracking, when it is directly linked to personal identifiers. As a result, a vast array of online market research activities—for example product preference polls that do not ask participants any personal information, or behind-the-screen tracking of user interactions in forums and online games—are not currently subject to regulation.[7]

Children's websites thus operate under very loosely defined, often highly self-serving, notions of "consent" that allow market researchers to collect vast databases of information on child users. For the academic researcher hoping to explore the ethical dimensions of online market research practices, however, this presents an immediate problem. The ethical standards required of academic research are established at various levels of governance, depending on the study's context and funding sources. Generally, however, these standards are much more stringent and comprehensive than those required of the private (marketing research) industry. The disjuncture between industry and

---

[7] One game-site that is particularly active in conducting marketing research, Neopets.com, flaunts this fact in their press kit, noting, "Neopets has the largest COPPA compliant (Children's Online Privacy Protection Act) online market research panel in the world, containing more than 50,000 12 and younger panellists complete with written parental permission" (Neopets Press Kit FAQ, 1999-2006).

academia can make it difficult to obtain ethical clearance to study marketing research processes and findings. For example, in academic research, informed consent and other considerations are necessary whenever human subjects are involved, whether the information gathered is "personally identifiable" or not. Market researchers, on the other hand, employ a much lower benchmark of consent and do not regulate the collection of non-personally identifiable data. Studies of marketing research processes and findings thus run a very real risk of coming across data that fail to meet academic ethical standards, and produce research without proper participant consent.

The question of consent is particularly crucial to ensuring ethical research. For example, in the case of the Office of Research Ethics (ORE) at Simon Fraser University, "expressed opt-in," one form of informed consent, is defined as "voluntary, informed, unambiguous, obtained before beginning the research and may be withdrawn at any time, and unless there is explicit consent at the time of data collection, there will be…no further analysis of the data initially collected" (*Policy R20.01* 1992: Requirement 8a). In order to qualify as "informed consent," the researcher must provide detailed information about the study's methods and purpose. A thorough description of the type of data to be collected is required, along with an accurate explanation of how it will be used. Although unanticipated secondary usage of findings is allowed as long as the findings are kept anonymous and published in aggregate form, the fact that informed consent was granted at the time of the initial data-collection distinguishes this exception from the marketing research practices explored above. In studies involving minors (defined as persons under the age of majority), parental consent is required, along with additional precautions and heightened ethical sensitivity.

Publicly-funded research involving children in Canada must also conform to special ethical standards. The *Tri-Council Policy*[8] (1998) includes children in its category of "vulnerable persons" warranting high ethical obligations. Not only must researchers obtain the "free and informed" consent of an authorized third party (parent or guardian), but they must also seek the consent of the subjects themselves:

---

[8] The *Tri-Council Policy* provides research guidelines for the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council of Canada (SSHRC).

Many individuals who are not legally competent are still able to express their wishes in a meaningful way, even if such expression may not fulfill the requirements for free and informed consent. Prospective subjects may thus be capable of verbally or physically assenting to, or dissenting from, participation in research. Those who may be capable of assent or dissent include: (a) those whose competence is in the process of development, such as children whose capacity for judgment and self-direction is maturing… (Article 2.7)

The notion of pairing parental consent with children's assent to participate is also found in the *Ethical Guidelines* of the Association of Internet Researchers (AoIR), which recommends that researchers adopt the broad consideration that "the greater the vulnerability of the author/subject—the greater the obligation of the researcher to protect the author/subject" (Ess & AoIR, 2002, n.p.). These standards not only go beyond the regulatory frameworks provided by government policies and industry guidelines, but they also give stronger consideration to children's participatory rights than is required by certain national and state laws.[9]

In comparing the standards applied within the private marketing research industry to those required of academic (or other publicly funded) research, it is thus clear that numerous points of contention exist between these two areas of research. This creates a special set of challenges for critical examinations of marketing research practices, not the least of which involves sufficiently reconciling two contrasting sets of ethical standards in order to secure ethical clearance. For the TOS study described above, the ethical approval process itself becomes an interesting case study in the disjuncture between industry and academia, and the problems this disjuncture can cause for academic research. It also illustrates significant oversights within current academic understandings of online research that will need to be remedied if children's rights as research subjects and cultural producers are to be respected within the digital context.

Case Study: Securing Ethical Clearance

---

[9] For example, in the US, children are not considered legally capable of giving informed consent (Ess & AoIR 2002).

The study of children's websites and TOS contracts described above was subject to three sets of ethical standards: Simon Fraser University's ORE *Policy*, the *Tri-Council Policy*, and the AoIR's *Ethical Guidelines*. Yet, the research subject—online marketing research—employed its own research design which did *not* meet the ethical standards required by any one of these policies. For example, both *Neopets.com* and *Toontown Online* incorporated child-generated content which could be used for market research, and both had secured a form of consent from the children and their parents to use that content in sweeping and ambiguous ways. On the other hand, neither of the sites provided any details about how the content would be used, and it was unclear that the children or parents understood that submissions could be used for market research purposes. It was thus unlikely that the consent the parents and children had granted to these sites qualified as "informed consent," and thus unlikely that these submissions would meet the level of consent required for inclusion in an academic study. In light of these conflicting interpretations of consent, the project was submitted for ORE review in early April 2005.

A series of questions was released by the director of the ORE, Hal Weinberg, requesting clarification on a number of aspects of the research design. Two of the questions involved the issue of whether or not the study would require the approval of the website owners/operators. Another related more specifically to the issue of participant consent:

> When parents give consent for their children to use the web site do they know that the site may be used for studies of the kind you will be conducting? (Weinberg, 2005a, n.p.)

In response, it was clarified that the majority of the contents under study were online submissions and "after-the-fact" communications[10] (such as pre-existing forum discussions). Since the websites were published online, it was argued that their contents fell within the scope of public domain, and thus permission from the site owners was unnecessary. This position was further supported by the commercial nature of the sites

---

[10] The only exception was in *Toontown Online*, Disney's MMORPG, where I would be privy to a restricted form of user interaction wherein users communicate by choosing from a limited selection of pre-defined communicative statements.

and the high amounts of publicity they generated, which relates to the AoIR guideline, "[T]he greater the acknowledged publicity of the venue, the less obligation there may be to protect individual privacy, confidentiality, right to informed consent" (Ess & AoIR, 2002, n.p.). For the final question concerning parental consent, excerpts from the TOS contracts were provided to show that by granting consent for their children (under the age of 13 years) to participate in the sites, parents[11] had already given explicit consent to abide by the website privacy policies and TOS contracts, which included transfer of intellectual property ownership of any child-generated content over to the site owners. Since the websites then published these "copyrighted" materials online and within the public domain, there was no legal reason why the submissions should not then become available for academic analysis.

On June 13, 2005, the project was categorized as Minimal Risk and provisionally approved by the Director, to be ratified by the Research Ethics Board (REB) at their next meeting (Weinberg 2005b). When the REB met in July, however, similar questions were raised, and data collection was suspended while legal opinion was sought. An email was forwarded to the Associate VP for Legal Affairs with a number of queries about the ownership of the data. The Board also expressed concern about two issues that are particularly relevant and significant to future work in this area:

> Does the web site policy sufficiently identify the parent/guardian understanding regarding the use of their child's data?

and

> Is there a Canadian law prohibiting the transfer of ownership of data given to the web site? (Weinberg, 2005c, n.p.)

On July 29, a letter (July 29, 2005) was sent by the REB announcing that in the opinion of the university's legal counsel the data was in the public domain, and that therefore the study had finally been granted ethical clearance. The disjuncture between marketing research ethics and academic ethical standards was ultimately resolved as a public

---

[11] Assuming that the child has not bypassed the system by providing false parent contact information.

domain issue. Nonetheless, the questions that had been put forth by Weinberg and the REB reveal that there remain a number of significant ethical dimensions of online research involving children that have not yet been adequately addressed within academic research ethics policies. It seems clear from both the case study and the content analysis research that the ethical standards for conducting research involving child-generated content online need to be re-examined within both industry and academia.

Discussion

From the above analysis, a number of preliminary recommendations for how ethics policies might be revised to better accommodate the special needs and vulnerabilities of children within the online context can be identified. The issue of informed consent cuts deeply into both questions of ownership as well as notions of the public domain. From the previous research in this area, it is unlikely safe to assume that the majority of parents or children are aware of the full implications of submitting content online. Rather than let commercial researchers diminish standards for informed consent by driving this content into the public domain, a more proactive approach must be undertaken by academic researchers to ensure that subjects' rights receive better protection within the online environment. This applies to both academic research practices as well as any unethical practices uncovered during the course of research. Furthermore, the requirement that children's assent be secured along with parental consent should be established within the university's ethical guidelines, a consideration that upholds the spirit of children's participatory rights found within the United Nations Convention on the Rights of the Child, and is in line with the recommendations of both the *Tri-Council Policy* and the *AoIR Ethical Guidelines*.

Until now, studies of documents and materials in the public domain have not been subject to ethical review, but then again until now contributing to the public domain was a predominantly adult activity. It has not been established that children truly understand the full meaning and implications of the public domain, or how this relates to their contribution to online culture. The apparent disconnect uncovered in the literature, between children's understanding of privacy (Turow, 2001; Shade et al., 2004) and the value they place on the Internet as a provider of "private spaces" (Livingstone, 2006) is

perhaps reflective of the high levels of ambiguity created by commercial sites around these issues. In the current climate, the task of educating youth about important subjects like privacy, legal contracts and authorship is often left to commercial websites run by toy companies and marketing researchers. Documents such as privacy policies and TOS contracts are difficult for children to understand and often omit important information about the business and legal relationships they seek to establish. By giving special consideration to child-generated content, we can reconfigure our approach to public domain and copyright issues to be more inclusive of children's special needs and vulnerabilities, while concurrently supporting their participation in this important realm of cultural life.

A good starting point has already been formulated within the *AoIR Ethical Guidelines*, which challenges the appropriateness of approaching child-generated content, such as websites, weblogs, and other online contributions, as part of the public domain. The questions raised by the AoIR (Ess & AoIR, 2002) document include:

> [A]re web pages created by minors—but often without much understanding of the possible harms some kinds of posted information might bring either to the author and/or others—to be treated as the same sort of document as authored by adults, who (presumably) are better informed about and sensitive to the dangers of posting personal information on the Web? Or are researchers rather required to exercise greater care in protecting the identity of minors - perhaps even to inform them when their materials may pose risks to themselves and/or others. (n.p.)

These considerations have led the AoIR to recommend that researchers apply the broad consideration that "[T]he greater the vulnerability of the author/subject—the greater the obligation of the researcher to protect the author/subject" in research involving documents and content created by minors[12] (Ess & AoIR 2002, n.p.).

---

[12] It is important to note, however, that it is also within commercialized websites that we find many of the best tools and richest opportunities for cultural participation currently available to children online. It is also one of the only areas where children's cultural contributions are valued, even while viewed in primarily proprietary terms. Conversely, regulation that restricts and excludes children's online access, as seen in the recent *Deleting Online Predators Act* (DOPA) in the US, takes away their rights instead of expanding them. It is important that opportunities for contribution are not eliminated, but that instead focus is placed on enhancing recognition and protection of children's participation in various spheres of social life.

The sweeping yet vague terms of the TOS contracts reviewed in the study do not provide sufficient details for the consent given by parents or children to truly qualify as "informed." Yet they nonetheless appear to comply with current regulatory and legal standards. It is clear from the CRTC's 1999 decision not to regulate the Internet that the Canadian regulatory system is out of touch with new technologies and the online practices of both industries and users. Meanwhile, the public domain resolution to the questions raised by Weinberg and the REB failed to address the underlying ethical issues involved. Upholding the websites' claims that children and their parents legally transfer ownership of their potential intellectual property to the sites by agreeing to the terms of service supports the misrepresentation that these minors' contracts are valid to begin with. While establishing the validity of TOS contracts is in the interests of marketing researchers attempting to package and sell data that is not legitimately theirs, it is not in the interest of child producers who may or may not understand the right of transfer (Cram et al., 1996). Until the regulatory and legal systems catch up to new digital realities, the responsibility for creating higher standards for online research rests in the hands of academic researchers.

Conclusion

The rise in prominence of digital technologies has in many ways aggravated the existing disjuncture between policy and practice, particularly where children are concerned. Ongoing debates about children and the internet focus on risks and opportunities that are defined primarily by adult concerns. Yet children have their own agendas when it comes to internet technologies, as well as their own thoughts and opinions about how the emerging digital culture should take shape. These aspects of children's technology use are causing deep transformations within contemporary childhood, including a move towards child-generated content and children's online communities. More than any other group of adults, the children's industries and marketing researchers have taken note of this shift, creating important online venues and opportunities that support children's increased cultural participation. Within these sites, the children's industries have created an ongoing dialogue with child users, much of which operates outside of existing regulatory frameworks. At the same time, however, marketing research practices have

now become more available for academic and public scrutiny than ever before. As seen in the case study and discussion above, the widening disconnect between industry and academic ethical standards can greatly complicate studies into online marketing research practices. Yet, rather than view this as an obstacle, academics with an interest in research ethics can also approach this disjuncture as a unique opportunity to enhance ethical standards and promote the rights of child participants within the online environment.

Researchers' obligations to child subjects must not end with mere compliance to existing standards and regulations. Oftentimes these do not adequately consider the growing relationship between children and commercial entities, nor do they support children's burgeoning rights as cultural producers. For this reason, the issues raised during this and future ethical clearance processes—of research projects involving both children's online communication as well as published texts and artefacts—should be systematically reviewed and incorporated into a new and evolving set of ethical guidelines and professional standards. Internet research and other forms of new media research are especially important to this process, as these studies often introduce new and unanticipated dimensions to emerging ethical questions. Accordingly, academics must find ways to study and critique unethical commercial research practices without conforming to the problematic norms they seek to establish. In so doing, academics can begin to construct a coordinated challenge to the lax commercial standards which currently set the status quo for the vast amount of online research that is secretly conducted on children.

References

Aikat, D. (2005). Click here for fun, games, friends…: Analyses of media content characteristics of children's web sites. Paper presented at the Annual Meeting of the International Communication Association, May 26-30, New York, NY.

Bennett, C. (2006). Keeping up with the kids. *Young Consumers: Insight and Ideas for Responsible Marketers, 7*(3), 28 – 32.

Burkell, J., Steeves, V., & Micheti, A. (2007). *Broken doors: Strategies for drafting privacy policies kids can understand*. Ottawa, ON: Office of the Privacy Commissioner of Canada.

Buying power of kids tops $18 billion. (2006, 9 May). *Press Release*. Packaged Facts. New York: PRNewswire. Retrieved October 26, 2006, from http://www.packagedfacts.com/corporate/press/Press_view.asp?Article=688

Canadian Marketer's Association. (2006) *Code of Ethics and Standards of Practice*. Ontario, Canada: Don Mills. Retrieved October 14, 2006, from http://www.the-cma.org/regulatory/codeofethics.cfm

Castronova, E. (2003) The right to play. Paper presented at the *State of Play Conference*, New York Law School, 13–14 November, URL (consulted January 2004): http://www.nyls.edu/docs/castronova.pdf

Children's Online Privacy Protection Act (COPPA). (1998). Washington, DC: Federal Trade Commission. Retrieved October 23, 2006, from http://www.coppa.org

Chung, G. & Grimes, S.M. (2005). Data Mining the Kids: Surveillance and Market Research in Children's Online Games. *Canadian Journal of Communication, 30*(4), 527-48.

Cook, D. T. (2000). The other child study: Figuring children as consumers in market research, 1910s—1990s. *The Sociological Quarterly, 41*(3), 487-507.

Coombe, R. J. (1998). *The cultural life of intellectual properties: Authorship, appropriation and the law*. Durham, NC: Duke University Press.

Cram, F., Ng , S.H. & Jhaveri, N. (1996).Young people's understanding of private and public ownership. In P. Lunt & A. Furnham (Eds.). *Economic socialization: The economic beliefs and behaviours of young people* (110-129). Cheltenham, UK: Edward Elgar Publishing Limited.

Dibbell, J. (2003). The Unreal estate boom. *Wired*, *11*(1), January. Retrieved January 8, 2004, from http://www.wired.com/wired/archive/11.01/gaming.html

Ess, C. & AoIR Ethics Working Committee. (2002). *Ethical decision-making and Internet research: Recommendations from the AoIR Ethics Working Committee*. Retrieved June 21, 2006, from http://www.aoir.org/reports/ethics.pdf

Frequently Asked Questions. (1999-2006). *Neopets.com Press Kit*. Neopets, Inc. Retrieved October 15, 2006 from http://info.neopets.com/presskit/press01.html

Furnham, A. (1996) The economic socialization of children. In P. Lunt and A. Furnham (Eds.). *Economic socialization: The economic beliefs and behaviours of young people* (11-34). Cheltenham, UK: Edward Elgar Publishing Limited.

Graft, K. (2006, 18 September). Is the core gamer a myth? *Next Generation -Interactive Entertainment Today*. Retrieved September 19, 2006, from http://www.nextgen.biz/index.php?option=com_content&task=view&id=3831&Itemid=2

Greenspan, R. (2003, October 9). Kids are media users too. *ClickZ.* Retrieved October 18, 2004, from http://www.clickz.com/stats/big_picture/demographics/article.php/5901_3089701

Grimes, S. M. (2006). Online multiplayer games: A virtual space for intellectual property debates? *New Media & Society, 8*(6), 969-990.

Grimes, S. M. (2007). Terms of service, terms of play in children's online gaming. In J.P. Williams & J. H. Smith (Eds.), *The player's realm: Studies on the culture of video games and gaming* (33-55). Jefferson, NC: McFarland Press.

Grimes, S. M. & Shade, L.R. (2005). Neopian economics of play: Children's cyberpets and online communities as immersive advertising in neopets.com. *International Journal of Media and Cultural Politics 1*(2), 181-98.

Helyar, R. (2007). Youth culture uncut: Youth tribes. *Young Consumers: Insight and Ideas for Responsible Marketers, 8*(2), 101-107.

Interactive Marketing Research Organization. (2000). *IMRO Code of Ethics*. Glastonbury, CT. Retrieved October 11, 2006 http://www.imro.org/profstds/code.cfm

Kapur, J. (1999). Out of control: Television and the transformation of childhood in late capitalism. In M. Kinder (Ed.). *Kids' media culture* (122-136). Durham: Duke University Press.

Kids account for one out of every five Internet surfers in the U.S.; More than 27 million American kids connect online. (2003, October 21). *News Release*. New York, NY:Nielsen/Netratings. Retrieved August 10, 2004, from http://www.nielsen-netratings.com/pr/pr_031021.pdf

Kids ages 2-11 lead growth in web page consumption, according to Nielsen/Netratings. (2004, November 17). *News Release*. New York, NY: Nielsen/Netratings. Retrieved July 4, 2007, from http://www.nielsen-netratings.com/pr/pr_041118.pdf

Klein, N. (2000). *No Logo: Taking aim at the brand bullies*. Toronto, ON: Alfred A. Knopf Canada.

Kline, S. (1993). *Out of the garden: Toys and children's culture in the age of TV marketing*. London, UK: Verso.

Kline, S. (2001). *Media use audit for B.C. teens*. SFU Media Lab: Burnaby, BC, Canada. Retrieved August 18, 2004, from http://www.sfu.ca/media-lab/research/report.html.

Lastowka, F. G. & Hunter, D. (2004). The laws of the virtual worlds. *California Law Review, 92*(1), 3-73.

Lindström, M. (2003). *BRANDchild: Remarkable insights into the minds of today's global kids and their relationships with brands*. Sterling, VA: Kogan Page.

Linn, S. (2004). *Consuming kids: the hostile takeover of childhood*. New York: New Press.

Livingstone, S. (2006). Children's privacy online: Experimenting with boundaries within and beyond the family. In Kraut, R. E., M. Brynin & S. Kiesler. (Eds.) *Computers, phones, and the internet: Domesticating information technology* (128-144). New York: Oxford University Press.

Loi sur la protection de la consommateur (Quebec Consumer Protection Act). (1971) Retrieved October 25, 2005, from http://www.consommation.info.gouv.qc.ca fr/sortie/fiche.asp?dossier=3412&sujet=75&miette=

Marketing Research Association. (2003). *The Code of Marketing Research Standards*. Glastonbury, CT. Retrieved October 14, 2006, from http://www.mra-net.org/pdf/expanded_code.pdf

Montgomery, K.C. (2000). Digital kids: The new on-line children's consumer culture. In D.G. Singer & J. Singer (Eds.), *Handbook of children and the media* (635-648). Thousand Oaks, CA: Sage Publications.

Moore, E. (2006). *It's child's play: Advergaming and the online marketing of food to children*. Washington, DC: Kaiser Family Foundation.

Nairn, A. (2006). Commercialisation of childhood? The ethics of research with primary school children. *International Journal of Market Research, 48*(2), 113-114.

Nearly 20 percent of the active online population are kids and teens, creating opportunities for marketers. (2002, August 13). *News Release*. New York, NY: Nielsen/Netratings. Retrieved October 18, 2004, from http://phx.corporate-ir.net/phoenix.zhtml?c=82037&p=irol-newsArticle&ID=538963&highlight=

Neuborne, E. (2001, August 13). For kids on the Web, it's an ad, ad, ad, ad world. *Business Week*, 108.

Personal Information Protection and Electronic Documents Act (PIPEDA). (2000). Ottawa, ON: Industry Canada. Retrieved October 3, 2006, from http://laws.justice.gc.ca/en/P-8.6/index.html

Policy R20.01. (1992, revised 2006). Burnaby BC: Office of Research Ethics, Simon Fraser University. Retrieved October 12, 2006, from http://www.sfu.ca/policies/research/r20-01.htm

Prescott, L.A. (2007, April 30). Virtual worlds ranking - Runescape #1. *Hitwise Intelligence.* Retrieved May 25, 2007 from http://weblogs.hitwise.com/leeann-prescott/2007/04/virtual_worlds_ranking_runesca.html

Preston C, (2004), Children's advertising: The ethics of economic socialisation. *International Journal of Consumer Studies, 28*(4), 364-370.

Quart, A. (2003). *Branded: the buying and selling of teenagers*. Cambridge, MA: Perseus Publishing.

Richtel, M. & Stone, B. (2007, June 6). For not-ready-for-MySpace girls, dress-up and diversions on the web. *New York Times*. Retrieved June 6, 2007, from http://www.nytimes.com

Rushkoff, Douglas (2006). *ScreenAgers: Lessons in chaos from digital kids*. Cresskill, NJ: Hampton Press.

Russo, J. (2001). How to read "terms of use" agreements. Presented at the *Computer Systems Laboratory Colloquium*, April 11. Stanford University: Palo Alto, CA.

Sandvig, C. (2000). The Internet disconnect in children's policy: A user study of outcomes for Internet access subsidies and content regulation. Presented to the *28th Conference on Communication, Information, and Internet Policy*. September 25, 2000. Alexandria, VA.

Seiter, E. (2004). The Internet playground. In J. Goldstein, D. Buckingham, & G. Brougére (Eds.), *Toys, games, and media* (93-108). Mahwah, NJ: Lawrence Erlbaum Associates.

Schor, J. B. (2004). *Born to buy: The commercialized child and the new consumer culture*. New York: Scribner.

Schor, J.B. & Ford, M. (2007). From tastes great to cool: Children's food marketing and the rise of the symbolic. *Journal of Law, Medicine & Ethics, 35*(1), 10-21.

Shade, L. R., Porter, N., & Sanchez Santiago, W. K. (2004). Everyday domestic Internet experiences of Canadian children and youth. Presented at *Digital Generations— Children, Young People and New Media*, July 26-29, 2004. Center for the Study of Children, Youth and Media, Institute of Education, University of London, London, U.K.

Smith, J. W., & Clurman, A. S. (1997). *Rocking the ages: The Yankelovich report on generational marketing*. New York, NY: HarperBusiness/HarperCollins Publishers, Inc.

Solomon, D. & Peters, J. (2005). Resolving issues in children's research. *Young consumers: Insight and ideas for responsible marketers, 7*(1), 68-73.

Steeves, V. (2006). It's not child's play: The online invasion of children's privacy. *University of Ottawa Law and Technology Journal, 3*(1), 169-88.

Steeves, V. & Kerr, I.A. (2005). Virtual playgrounds and buddybots: A data-minefield for tweens. *Canadian Journal of Law and Technology, 4*(2), 91-98.

Steeves, V. & Wing, C. (2005). *Young Canadians in a wired world phase II: Trends and recommendations*. Ottawa, ON: Media Awareness Network.

Sutherland, A. & Thompson, B. (2001). *Kidfluence: Why kids today mean business*. Toronto, ON: McGraw-Hill Ryerson Ltd.

Taylor, T.L. (2002). Whose game is this anyway?: Negotiating corporate ownership in a virtual world. In F. Mäyrä (Ed.), *Proceedings of Computer Games and Digital Cultures Conference.* Tampere, Finland: Tampere University Press. Retrieved August 28, 2003, from http://social.chass.ncsu.edu/~ttaylor/papers/Taylor-CGDC.pdf

Television advertising leads to unhealthy habits in children; Says APA Task Force. (2004, February 23). American Psychological Association. *APA Press Releases.* Retrieved November 23, 2005 from http://www.apa.org/releases/childrenads.html

Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (1998) Ottawa, ON: Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities

Research Council of Canada. Retrieved October 11, 2006, from http://www.pre.ethics.gc.ca/english/pdf/TCPS%20October%202005_E.pdf

Turow, J. (2001, March). *Privacy policies on children's websites: Do they play by the rules?* (Report Series). Annenberg Public Policy Centre of the University of Pennsylvania: Philadelphia, PA, USA.

Turow, J. (2003, June). *Americans and online privacy: The system is broken*. Philadelphia, PA: Annenberg Public Policy Centre of the University of Pennsylvania.

Weinberg, H. (2005a, 13 April) *Personal correspondence*. Burnaby, BC: Office of Research Ethics, Simon Fraser University.

Weinberg, H. (2005b, 13 June) *Personal correspondence*. Burnaby, BC: Office of Research Ethics, Simon Fraser University.

Weinberg, H. (2005c, 13 July) *Personal correspondence*. Burnaby, BC: Office of Research Ethics, Simon Fraser University.

Emerging Legal Issues in the Collection and Dissemination of Internet-Sourced Research Data: Part I, Basic Tort Law Issues and Negligence[1]

T.A. Lipinski, J.D., LL.M., Ph.D.
University of Wisconsin—Milwaukee
PO Box 413
Milwaukee, WI 53202
tlipinsk@uwm.edu

Abstract

This article is the first in a series exploring the legal basis for liability on the part of researchers who "observe" and "collect' data from online forums such as a listserv, discussion board, blog, chat room and other sorts of web or Internet-based postings. Determining whether a data subject is "harmed" in the process requires examining existing legal concepts relating to the law of personal injury, i.e., tort law. This article establishes the groundwork for future articles by applying the law of negligence to online research settings. Concepts such as duty of care, proximate and superseding cause, assumption of risk and contributory negligence are discussed and applied to online research data collection and dissemination. While there is some legal risk, the author concludes that under most circumstances a plaintiff would be unsuccessful in pursuing a legal claim against an individual researcher. Nonetheless as online research activity increases the litigious nature of our society underscores the need for researchers to be aware of how liability can arise and the factors involved in assessing that liability, the level of legal risk and appropriate risk-avoidance strategies.

Introduction

---

[1] Portions of this article represent an updated and expanded version of a discussion that first appeared in the Journal of Information Ethics (Lipinski, 2006).

This article is the first instalment in a series of four articles examining the possible basis for legal liability of researchers who use the Internet in the collection of research data. Specifically, it examines the potential legal issues associated with the protocols of ethnographers who use listserv, discussion board, blog, chat room and other sorts of web or Internet-based postings as the source of their data. The author assumes that the forum for participation is legitimate, in that the list, board, blog, chat, site, etc. is not created or otherwise concocted by the researcher, but involves situations akin to off-line practices where a researcher would be listening to the conversation or watching the behaviour of subjects in a public place. The forum is "legitimate" in this sense but the researcher is either reviewing the log, transcript or other record of participation made by others or is observing ("listening in on") the online conservation as it occurs among forum participants. In the latter case the researcher might be a hidden observer or merely an obvious but silent member, that is, "logged-in" as a user but choosing instead not to participate. In this case it could be said that the researcher is somewhat of an interloper. The legal implications (potential and likely consequences) of either asynchronous review or synchronous observation (data collection) and use (dissemination) of the information collected are discussed.

The author also assumes that the host location of the listserv, discussion board, blog, chat room or other sort of web or Internet-based forum is within the jurisdictional boundaries of the United States, thus, in this discussion, U.S. law applies. While the basic legal concepts may apply across international borders, researchers unsure of which laws apply should consult additional resources or make similar analyses within their specific context. Even within the United States, some legal concepts may vary across state lines (or federal jurisdictional boundaries such as judicial districts and circuits) and again reference to the context of the particular researcher may be necessary. As a result, this article offers readers an awareness of the variety of legal issues that might arise in the general context (data collection via observation) as well concrete examples in the selected specific context (online observations subject within U.S. jurisdictions). It offers the informed and experienced opinion of the author but should not be taken as legal advice. If legal advice is needed by the journal readership, the appropriate source for that advice should be sought. The opinions offered are based on interpretation and application of the

relevant precedent, but little if any case or statutory law exists in the precise ("on point" to use the appropriate legal phrase) context of data collection from online forums by researchers in this manner. To be sure, that is a good thing, but with the ever-increasing recourse of litigation to settle perceived harms or alleged disputes the law is sure to develop in this regard. It is the goal of the author that the analysis presented in this series offers the initial legal context in which such issues would be adjudicated and more importantly resolved.

This first article in the series discusses legal aspects of harm based in claims of negligence that arise from the online practices of researchers. The second article in the series discusses the claims against researchers based in defamation, either for comments made by the researcher or by a forum participant that the researcher then disseminates in other publications such as proceedings, articles, chapters, etc. The next instalment in the series reviews the potential for harm based on several theories of privacy invasion. A fourth and final article discusses copyright and online licensing. The latter concept includes assessment of the enforceability of terms and conditions governing access to and use of listserv, discussion board, blog, chat room and other web forums to which the researcher assents by use of a click-on or similar mechanism. This article does not discuss the use of the Internet as a collection device, e.g., online surveys. Since these practices do not involve direct interview or interaction with or treatment of human subjects much of the precedent involving negligence, informed consent and related issues is not relevant, however, it is reviewed initially for guidance in suggesting the applicable legal standards of conduct. While it is concluded that the likelihood for a success of such claims is small, analyses will include steps that researchers can take to ensure that such risk remains small.

Claims Based in Tort Law: General Concepts of Harm

In the United States, tort law is the law of private injury, of civil (as opposed to criminal and public) harm. (In other countries with Roman and civil law traditions the term "delict" may be used instead.) Black's Law Dictionary (2004) defines a "tort" as a "civil wrong, other than breach of contract, for which a remedy may be obtained, usually in the form of damages; a breach of a duty that the law imposes on persons who stand in a

particular relation to one another." In light of this definition the following question can be posed and refined: Is there any legal liability for researchers who "observe" research subjects, parallel to ethnographers that watch or listen to participant interaction in public places and without involvement of the researcher, i.e., prodding of the subjects to act or respond? More accurately, is there a potential for liability of researchers who read the script or log of comments posted by others (the subjects) in various Internet or web-based forums, then analyze that text, and publish summarizations or excerpts of that text in reports, articles, or other scholarly communications? Can it be said that a civil wrong occurs in this process? Said in another way, what legal duty, if any, is placed upon researchers in this stead?

The researcher may be reviewing the log, transcript, or other record of participation made by others or may be observing ("listening in on") the online conservation as it occurs among forum participants. In the latter case the researcher might be a hidden observer or merely an obvious but silent forum member, i.e., "logged-in" as a user but choosing instead not to participate. It goes without saying that online researchers typically have no ill-intent, but also that online researchers likewise do not enter into public web spaces under false pretense or disguise. In light of these factual assumptions, the legal analysis of either subsequent review or contemporaneous observation is the same.

As long as participants can be said to interact in a public forum, one open to anyone, without promise of discrimination by the host, the author considers the forum a public one. The host will not typically make such assertion or promise of discriminating access. Even where the forum is intended to be dedicated for some subset of the public, e.g., users interested in a particular hobby or activity, meeting some particular demographic or having some other characteristic, the host will make no guarantee to that effect. It may be obvious even to the reader untrained in law, but this is done for legal reasons, as such assertion or promise would raise the expectation on the part of the participant and this, as discussed below, might raise the legal duty of care owed to the forum participant by the forum host.

Forum participants do not have any enforceable rights against the host to enforce such interest *in* parameters or even behaviour guidelines. Even if registration, sign-up, or similar permission process must occur first, forum hosts will make no guarantee to a user that other users meet the parameters of that subset. What host could guarantee that no online predator or other third party with ill-intent would ever log into a forum designed for children, or that a site devoted to Seder practices and recipes would never be visited by neo-Nazi or white-supremacists? The closest analogous case demonstrating this principle is *Noah v. America Online, Inc.*, (2003, p. 534) where the plaintiff "claimed that the ISP wrongfully refused to prevent participants in an online chat room [devoted to Muslim topics: the 'Beliefs Islam' and the 'Koran' chat rooms] from posting or submitting harassing comments that blasphemed and defamed the plaintiff's Islamic religion and his co-religionists." In dismissing the case, the federal district court observed that the terms and conditions in the service agreement regarding online comportment entitled "Community Guidelines" created no contractual duty on the part of the service provider to "police" its space or ensure that only relevant participants enter a particular room; rather the "plain language of the Member Agreement makes clear that AOL is not obligated to take any action" (*Noah v. America Online, Inc.*, 2003, p. 545). Of course the focus of this article and the others in the series is upon the legal consequences of researcher conduct, not host conduct.

For liability to exist some legal injury must occur in the act of research or its aftermath. The most common cause of action for such injury would be to claim that the researcher is somehow negligent. In other words, the lawsuit would be based on the assertion that the researcher breached a duty of care owed to the subject (see discussion below) for the injury that results from either the observation alone or from dissemination of that observation to others. Commentators have concluded that the likelihood of suit against researchers is increasing: "Although the case law on human subject research is just beginning to emerge, plaintiffs should be able to successfully establish negligence claims against researchers by establishing a duty of care based on the special relationship between researchers and subjects" (Jansson, 2003, p. 262). However, these cases involve human subjects in clinical, medical, or related studies. Even within this genre of cases, situations where no physical contact is perpetrated have not met with success. An

example would be a claim based on the "injury" that results from no longer being a part of a study group: Expulsion occurs much to the chagrin of the subject, as the subject desires to continue taking the experimental drug, receiving the experimental treatment, and the subject believes that their health is declining as a result of no longer being allowed to participate (*Spenceley v. M.D. Anderson Cancer Center*, 1996). While it is possible to craft a claim for non-contact harms, e.g., injury from either intentional or negligent "infliction of emotional distress," in those jurisdictions that recognize such claims the claim must nonetheless be "linked to a physical injury or the threat of one" (Morreim, 2003, p. 70). The question is whether any injury, other than a specific act such as a violation of privacy (discussed in a subsequent article) or defamation (discussed below) for re-publication by the researcher of a statement made by a participant, is perpetrated by merely observing or listening to another's conversation or comment when that expression is text-based and made in an open forum on the Internet, or whether any injury results from the dissemination of that conversation or comment. This "observation" does not involve an invasive act into the physical space of the participant. This is opposed to contact-based "injury," while not causing harm can nonetheless be said to "injure" a participant.

There have been cases against medical groups where the patient-participant suffers no physical harm (the operation was indeed successful) but is still considered to be injured when the medical group substitutes another surgeon for the preferred one, a so-called ghost doctor. Such circumstances can be the subject of a claim for battery (*Perna v. Pirozzi*, 1983); *Pugsley v. Privette*, 1980); *Guebard v. Jabaay*, 1983). However the injury is not the result of physical harm *per se*, of contact gone awry, but from an intentional harm, the purposeful substitution of one doctor for another: "An act which, directly or indirectly, is the legal cause of a harmful contact with another's person makes the actor liable to the other, if . . . the contact is not consented to by the other or the other's consent thereto is procured by fraud or duress" (Restatement (First) of Torts 1934, § 13).

Understanding the Elements of Negligence

The circumstance would be rare where the online researcher could be said to intend for harm to result from the observation of listserv, discussion board, blog, chat room and other web forum participation or the dissemination of that observation through publication, i.e., in the form of a proceeding, article, book chapter, pr other publication. It is more likely the circumstance that the claimed harm occurs, to use the non-legal but no less accurate conceptualization, by accident. Nonetheless, the forum participant still cries foul believing some compensable injury occurs as a result of the observation or dissemination. In the law of tort, of injury, the general label assigned such non-intentional harms is negligence. Understanding the law of negligence and its application to the practices of online researchers identifies the legal basis on which such a claim would lie and the likely result a court would render in response to such claim. Again the reader is cautioned that the law of negligence can vary from jurisdiction to jurisdiction. This article discusses general concepts and offers example of selected cases. However, the concept of negligence does exist within all jurisdictions of the United States and in other countries so the considerations discussed here would be relevant in multiple environments.

Negligence is "conduct which falls below the standard established by law for the protection of others against unreasonable risk of harm. It does not include conduct recklessly disregardful of an interest of others" (Restatement (Second) of Torts, (1965, § 282). To prevail in a claim of negligence against another party the forum participant plaintiff must establish that the person who caused the harm, called the *tortfeasor*: 1) owed a duty of care of to the plaintiff, 2) there was a breach of that duty, 3) the breach was the proximate (or legal) cause of the harm, and 4) and harm (identifiable and measurable) results that is "legally compensable" (Lee, 2005, § 3:2). The lynchpin of establishing a legal duty is *foreseeability*. According to Black's Law Dictionary (2004) the resulting harm must have been "reasonably anticipatable" before the defendant would have a duty to protect against its occurrence. The concept reflects a logical response by the courts: In general it is proper to make a person responsible for the harm one causes, but where it is not reasonable to anticipate harm the legal system will conclude it is likewise not reasonable to expect that the harm can be prevented. In other words, the duty of care to protect others from harm extends to those harms which are foreseeable. Where

harm is not foreseeable one cannot be expected to exercise a duty to protect others from it. The failure of that prevention is the breach, the second element.

Duty of Care and Foreseeability

While it might be tempting to believe that all accidents are preventable and therefore foreseeable in the eyes of the law, the law does not see it that way. Foreseeability is a legal concept and a determination of whether a duty of care exists is a question of law, i.e., it is a decision for the court not the jury to make. The Restatement (Second) of Torts (1965, § 435, emphasis added) offers two considerations regarding the concept of "Foreseeability of Harm or Manner of its Occurrence": (1) If the actor's conduct is a *substantial factor* in bringing about harm to another, the fact that the actor neither foresaw nor should have foreseen the extent of the harm or the manner in which it occurred does not prevent him from being liable. 2) The actor's conduct may be held *not* to be a *legal cause* of harm to another where after the event and looking back from the harm to the actor's negligent conduct, it appears to the court highly extraordinary that it should have brought about the harm." Foreseeability is but one of several factors courts use in determining whether a legal duty exists.

Cases where the plaintiff's claim was based upon negligent dissemination of content such as entertainment media (video games) or the printed word (books, magazines, etc.) are instructive. The defendants in these cases engaged in speech that remained unprotected under the First Amendment, in that the dissemination of the content that was not defamatory or inciting. In *Sanders v. Acclaim Entertainment, Inc.* (2002, 1271), the Columbine High School shooting case in Colorado the district court weighed four factors in its duty of care analysis: "1) foreseeability of the injury or harm that occurred; 2) the social utility of Defendants' conduct; 3) the magnitude of the burden of guarding against the injury or harm; and 4) the consequences of placing the burden on the Defendants." Other articulations are possible. Consider the extended list of factors the court used in *Juarez v. Boy Scouts of America, Inc.* (2000, p. 29, citing *Rowland v. Christian*, 1968, p. 113): "These factors include: (1) the foreseeability of harm to the injured party; (2) the degree of certainty that the injured party suffered harm; (3) the closeness of the connection between the defendant's conduct and the injury suffered; (4)

the moral blame attached to the defendant's conduct; (5) the policy of preventing future harm; (6) the extent of the burden to the defendant; and (7) the consequences to the community of imposing a duty to exercise care, with resulting potential liability." Applying these factors the court concluded that the "*Rowland* factors supports the imposition of a duty of care on the Scouts to have taken reasonable protective measures to protect Juarez from the risk of sexual abuse by adult volunteers involved in scouting programs, such as warning, training or educating him (either directly or through his parent or adult volunteers) about how to avoid such a risk. Of these factors, only the lack of moral blame is determined in favor of the Scouts. The rest are unequivocally in favor of Juarez and the existence of a duty" (*Juarez v. Boy Scouts of America, Inc.*, 2000, p. 35). Whatever set of factors are used, the duty of care concept retains one overreaching goal, to determine the legal point of responsibility between the plaintiff's lack of forbearance and the defendant's lack of caution. The factors weigh the cost and benefit of excusing the conduct of the defendant or placing responsibility upon the defendant for risk avoidance instead of with the plaintiff.

The second and third elements of the *Sanders* duty of care articulation inquires, to use the old adage, whether an ounce of prevention can result in a pound of cure or whether the opposite is true, that the burden to prevent the occasional and unanticipated injury would far outweigh the gain of an additional safe circumstance. In *James v. Meow Media*, (2002, p. 693, emphasis added), the Heath High School shooting case in Paducah, Kentucky, the court observed: "We almost appeared to say [in *Watters v. TSR, Inc.* (1990, p. 381), the *Dungeons and Dragons* suicide case] that the *costs* of acquiring such knowledge [burden on producer to identify which consumers might have a propensity for extreme dysfunctional responses] would so outweigh the *social benefits* that it would not be negligent to abstain from such an investigation. We can put the *foreseeability* point a little more precisely, however. It appears simply impossible to predict that these games, movies, and internet sites (alone, or in what combinations) would incite a young person to violence." In contrast, and in a rare case, where the risk is more likely (explicit advertisement in a national magazine) and the harm greater (contract murder), the court found a duty of care present involving the negligent publication of a "Gun-For-Hire" advertisement: "The evidence in this case fully supports the jury's determination that

defendants did not conform to the applicable standard of conduct... A risk becomes unreasonable when its magnitude outweighs the social utility of the act or omission that creates the risk" (*Braun v. Soldier of Fortune Magazine, Inc.*, 1992, pp. 1328-1329). Unlike a video game, the Soldier of Fortune advertisement if responded to by a reader and taken to a logical conclusion exhibits more than a remote possibility that severe harm will ensue, such is not the case with playing a video game. Said in another way, many lawful and non-injurious incidents can result from playing video games; shooting rampages are the exception, but unlawful and injurious incidents are far more certain to result from making a serious inquiry to a "Gun-For-Hire" advertisement in a magazine for professional soldiers *(Norwood v. Soldier of Fortune Magazine, Inc.*, 1987).

The fourth element assesses the potential fallout against the defendant and like-situated defendants in the future. In cases involving negligent speech, courts may assess the impact on future speech, to determine whether a so-called chilling effect occurs. Typically a concept associated with constitutional (state as opposed to private actions) questions of free speech relating to government regulation of citizen speech. Black's Law Dictionary (2004), indicates that the concept is also used "broadly, the result when any practice is discouraged." Thus even in civil actions where the claimed injury results from speech, courts often consider the First Amendment. For example in *Smith v. Linn* (1989, p. 127), the court concluded that the free speech rights of the publisher protected it from liability for the death of a person who followed the regimen outlined in The Last Chance Diet: "As appellant has not persuaded this Court by any existing case law that its cause of action against Lyle Stuart, Inc., can withstand judicial scrutiny, in view of the first amendment right of the publisher to publish the diet book in question, we find no error on the part of the trial court in determining there was no genuine issue of material fact necessitating trial and in accordingly entering summary judgment." The dissemination of the online observation made by researchers as part of their scholarship in the form of proceedings, articles, chapters, etc. is protected speech, e.g., is not defamatory or inciting. Imposing a duty of care for possible harm from that speech, could be said to chill the progress of such speech in the future. This would have a negative impact on the production of future scholarship. As a result the social utility of the imposition of a duty of care would be reduced instead of increased. As a result, there may be more powerful

factors motivating a court to conclude a lack of duty exits in the context of online observation and dissemination than in the production and distribution of video games. This is not to assert that video games are without social value, rather a court may view scholarly research as a more valuable social product than entertainment products such as video games, or at least, as valuable.

A rare case where a court concluded that a newspaper was responsible for negligent publication of a person's name and address involved unique circumstances, namely, the danger that might arise when the name of a crime victim is published. In *Hyde v. City of Columbia*, the victim though escaping from a first abduction by her assailant was "terrorized on seven different occasions" after the newspaper published a report on the incident along with personal information that the assailant used to identify and locate her (1982, p. 253). The *Hyde v. City of Columbia* (1982, p. 269) court concluded "that the name and address of an abduction witness who can identify an assailant still at large before arrest is a matter of such trivial public concern compared with the high probability of risk to the victim by their publication, that a news medium owes a duty in such circumstances to use reasonable care not to give likely occasion for a third party (assailant still at large) to do injury to the plaintiff by the publication." Intimidation of witnesses might also make them less forthcoming which in turn might jeopardize the progress of the criminal justice system; this results in less social utility as well. The court's comment again reflects the balancing of cost and benefit that occurs in the assessment of duty. Unlike the circumstances of *Hyde* the fourth factor may be more relevant in the case of online researchers or any researchers for that matter, as placing the burden for anticipating all possible harms upon individual researchers may have a deleterious impact on future practice and the impact would far outweigh the social benefit of protecting from harm forum participants who choose to interact in public settings. In regards to an action for negligent publication, the plaintiff is not asserting that the information is false or inaccurate, but that its truthfulness and dissemination of that "truth" leads to injury. If the basis for the claim is injury from falsity, the likely cause of action would lie in defamation (discussed below).

Proximate Cause and Foreseeability

Even if a duty of care is established, evidence of a breach of that resulting in identifiable and measurable harm a court may still rule that the conduct of the researcher is not the legal or proximate cause of the harm suffered. The concept of proximate cause often involves a policy analysis that too can focus on foreseeability, as "foreseeability along with actual causation, is an element of proximate cause in tort law" (Black's Law Dictionary, 2004). Determinations of breach and causation are questions of fact not law and thus are questions for the trier of fact, which could be the judge or could be the jury in a jury trial. (In the discussion that follows the word "court" includes jury.) However, issues of causation before a jury must still be informed by the law of causation or proximate cause. Courts contemplate the present desirability as well as impact on future alleged tortfeasors when concluding that the breach of the foreseeable harm was in fact *the* (legal) cause of the plaintiff's injury. In other words, is the chain of events close enough in terms of a logical connectedness that it should form the basis of legal responsibility, or is the harm at the finish of the event-chain too remote to contemplate? Knock the first domino down and one can clearly see how each successive tip of the block hits the next and causes the next to tip and so on, but should we hold the person who knocked the first block down by accident responsible for the fall of the last? The more intricate the chain, even if the dots can be readily connected, the less likely a remote cause will be found.

Interceding in this legal backtrack of events is the concept of superseding cause or the acts of the plaintiff or third party that contribute to the harm which can also operate to frustrate the plaintiff's claim. According to the Restatement (Second) of Torts (1965, § 440, emphasis added), "Superseding Cause Defined": A superseding cause is an act of a third person or other force which by its *intervention* prevents the actor from being liable for harm to another which his *antecedent negligence is a substantial factorSanders* case (*Sanders v. Acclaim Entertainment, Inc.*, 2002, p. 1276), the court identified the elements of a superseding cause: "A superseding cause exists when: 1) an extraordinary and unforeseeable act intervenes between a defendant's original tortious act and the injury or harm sustained by plaintiffs and inflicted by a third party; and 2) the original tortious act is itself capable of bringing about the injury." in bringing about… A superseding cause relieves the actor from liability, irrespective of whether his antecedent negligence was or

was not a substantial factor in bringing about the harm. Therefore, if in looking back from the harm and tracing the sequence of events by which it was produced, it is found that a superseding cause has operated, there is no need of determining whether the actor's antecedent conduct was or was not a substantial factor in bringing about the harm."

Did the intervening criminal acts of the Dylan Klebold and Eric Harris supersede any fault on the part of the video game manufacturers? Foreseeability, superseding cause, and proximate cause concepts are present and interrelated in the third element of negligence. In other words, does the video game manufacturer have a responsibility to take into account the fact that some individual somewhere might engage in criminal activity reminiscent of a scene or play in the video game? No. "Just as *foreseeability* is central to finding that a *duty* is owed, it is also 'the touchstone of *proximate cause*' and of the superseding cause doctrine. Moreover, a superseding cause relieves the original actor of liability when 'the harm is *intentionally caused* by a third person and is *not within the scope of the risk* created by the actor's conduct'" (*Sanders v. Acclaim Entertainment, Inc.*, 2002, p. 1276, emphasis added, quoting *Webb v. Desert Seed Co.* (1986, pp. 1062-1063), quoting Restatement (Second) of Torts (1965, § 442B). Likewise the court in *James* concluded that an intentional criminal act breaks the chain of causation, such that no further dominos fall: "Even if this court were to find that the defendants owed a *duty* to protect James, Steger, and Hadley from Carneal's violent actions, the plaintiffs likely have not alleged sufficient facts to establish the third element of a *prima facie* tort case: proximate causation. The defendants argue that even if they were negligent, Carneal's *intentional, violent actions constitute a* superseding cause of the plaintiffs' damages and sever the defendants liability for the deaths of Carneal's victims. Generally, a third party's criminal action that directly causes all of the damages will break the chain of causation" (*Sanders v. Acclaim Entertainment, Inc.*, 2002, p. 699, emphasis added). Criminal acts as well as other intentional acts are too remote to contemplate and courts will not conclude that a person at the start of a chain of events is responsible for such end harm.

Assessment of Negligence for Researchers in Online Settings

Consider the following facts. A researcher who studies adolescent sexuality collects data (postings) from an online forum dedicated to some subset of teenagers, i.e., those with a particular interest or characteristic. From analysis of the comments, the researcher concludes that a number of the participants are either likely to engage in early sexual promiscuity or prone to sexual exploitation by others. The researcher disseminates the findings in a scholarly journal. A sexual predator comes across the journal article online, in which the sources of the data (the forums) are listed in the methodology section of the paper. The predator later visits the sites, engages in numerous chat sessions and convinces several of the female teenage participants to meet him for an amorous rendezvous. He later sexually assaults several of them. The predator is arrested and prosecuted. During investigation and trial it comes out that the predator selected his victims based on the vulnerability profiled by the researcher and from the same chat rooms cited by the researcher in the article. The families of several of the victims sue the researcher and publisher of the journal. Is the researcher in anyway legally responsible for the harm that befell the forum participants?

Applying the concepts of negligence to the scenarios of researchers who observe subjects in online forums and disseminate some aspect of that observation in the context of scholarship suggests that a claim of negligence is unlikely to be successful. In clinical or other human subject research, establishing a duty to inform or protect the subject from undue risk is somewhat obvious, but what duty of care do researchers merely "listening in" on online public forum conversations have when snippets of the "conversations" are repeated to others in the context of scholarly publications? What additional risk is there that the subject by choosing to participate in the forum did not already assume in the first instance? These questions raise concepts that are embedded in the tort law.

Assumption of Risk and Contributory Negligence for Dangerous Behaviour

Forum participants may be said to "assume a risk" of harm for the consequences of posting identifying information or images, expressing personal opinion, willingness to agree to off-line contact, et cetera. Assumption of risk is a legal concept and can operate to absolve the defendant of liability in cases where a defendant acted in negligence. The

Restatement (Second) of Torts (1965, § 496A) "Assumption of Risk" "General Principle" states that "a plaintiff who voluntarily assumes a risk of harm arising from the negligent or reckless conduct of the defendant cannot recover for such harm." In the predator chatroom scenario, it can be said that the female forum participants assumed the risk of the subsequent assault by engaging in continued contact with the predator and agreeing to meet with the stranger with an understanding that a tryst was intended. Do not, however, confuse this tort law concept with criminal law concepts of culpability. As far as the criminal law is concerned, the victims are not responsible for their assault. In addition, if any of the teenagers were minors, legal issues of consent may also be involved. Under the criminal law of some states a minor or a minor under a certain age cannot legally consent to sexual relations.

A forum participant can also "contribute" to the negligent environment. In general, when the plaintiff is more negligent than the defendant, the plaintiff cannot recover for the injury. Where the plaintiff is somewhat negligent but not more than the defendant, most jurisdictions will reduce the plaintiff's recovery by an amount representing, in percentage points, the contributing negligence of the plaintiff. "The same conduct on the part of the plaintiff may thus amount to both assumption of risk and contributory negligence, and may subject him to both defenses… assumption of risk rests upon the voluntary consent of the plaintiff to encounter the risk and take his chances, while contributory negligence rests upon his failure to exercise the care of a reasonable man for his own protection… A subjective standard is applied to assumption of risk, in determining whether the plaintiff knows, understands, and appreciates the risk. An objective standard is applied to contributory negligence" (Restatement (Second) of Torts (1965, §496A, comment d).

The age of the plaintiff can affect the ability to appreciate and therefore assume the risk. "If by reason of age, or lack of information, experience, intelligence, or judgment, the plaintiff does not understand the risk involved in a known situation, he will not be taken to assume the risk, although it may still be found that his conduct is contributory negligence because it does not conform to the community standard of the reasonable man" (Restatement (Second) of Torts (1965, § 496D, comment c, "Knowledge And Appreciation Of Risk"). In the hypothetical predator chat room

scenario, the teenage victims would likely be deemed mature enough to understand and assume, in terms of the civil tort law, the risk of ignoring the charge drummed into children from an early age to "never talk to strangers" and "never get into a car with a stranger" and its modern incarnation against entering into relationships with chat room participants, especially where chat of sexual activity is involved. Short of assuming the risk of such trysts, the teenagers in terms of the civil tort law contributed to their harm.

Both concepts of assumption of risk and contributory negligence may operate to insulate researchers from any risk of liability from observation and dissemination of information others voluntarily submit to online forums. Before these concepts come into play, however, a court would first need to determine that the researcher was in some way negligent. Is this a realistic possibility?

Making the Claim of Negligence, Elements Applied to the Circumstances

For such claim to be successful it would be as if the plaintiff (the research subject) charges that because you (the researcher) do not know me or did not ask my consent you should not be listening to my conversation or watching my actions and recounting what transpires, even if those occur in a public place. While we were all taught that it is rude to listen to others' personal conversation and that it is not polite to stare, the law would hardly find a remedy for such an imagined harm. So what legal duty of care does a researcher have? Can it be said that there is a duty not to listen to or observe the participation of others in a public place or, in this instance a virtual space, an online forum open to all comers? Is there a duty to cease observation when the opportunity arises to do so, to avert your online eyes or shun your online ears so to speak or to refrain from telling others what you observed through subsequent dissemination? Ridiculous! In fact, it could be argued that such forums are more open than traditional public spaces in the real world, as the physical casual observer or passer-by might by accident overhear part of a conversation or cast a distracting glance, the observer might need to take obvious and affirmative acts to maintain proximity in order to continue listening to or maintain visual contact of a particular subject. Yet, in the online world, the purpose and expectation of such forums is to allow all comers to read one's thoughts or opinions, or view one's art work, or cell-phone self portrait, or myspace.com video autobiography,

with full understanding of the likelihood that others will in return make a similar response, comment, critique, etc.

A court would not likely conclude that there is a duty of care to refrain from observing or recording the communication of participants on a listserv, discussion board, blog, chat room and other sort of web or Internet-based posting. Dissemination of the communication would fall under a similar rule. This assumes that the researcher does not disseminate identity information but follows an anonymity protocol, thus avoiding the circumstances of *Hyde v. City of Columbia*. Recall that the dissemination of identity information in that case was that of a crime victim in the context of a still at-large assailant. Also recall also the facts of the hypothetical and the concept of foreseeability. Is it possible that online sexual predators will use the Internet to identify, seek out, and pursue victims on various online forums? Is it also possible that predators may be alerted to the Internet whereabouts of these potential forums by third parties, e.g., advertisements on and off line promoting new chat rooms dedicated to teenage topics, an article in a local newspaper about the proliferation of online dating services or a researcher who studies online communities and publishes on the topic? Yes, but should the law place the legal burden of foreseeability for the clandestine activities of predators on those third parties, holding those third parties to a legal duty of care to ensure that such harm does not occur from the dissemination? It is unlikely that a court would hold a third party speaker or publisher to this standard.

If identity information is further disseminated from the initial posting but without the immediacy of a recent and known injury such as the first assault on the victim in *Hyde*, a court may still conclude that foreseeability is absent. Where the dissemination includes identity information the case could fall under the *Braun v. Soldier of Fortune Magazine, Inc.* holding. However, the unique circumstances of that case, i.e., Gun-For-Hire advertisements involving the solicitation of dangerous criminal activity, limit its application to the unusual scenario. As discussed above, the concept of duty of care includes various shifting scales of evaluation. Foreseeability may be present in situations where the severity of potential harm is increased and is coupled with a parallel increase in the likelihood that harm the will occur, especially where the dissemination poses a direct and unique link to the harm. This analysis would by logic apply to the dissemination of

threats of serious physical harm, hate, or otherwise harassing speech. In the case of *Braun v. Soldier of Fortune Magazine, Inc.* it was the publication of contact information in the context of the advertisement that evidenced a foreseeable outcome. The original advertisement that appeared in the June 1985 through March 1986 issues of *Soldier of Fortune* included the phone and address information of the Gun-For-Hire solicitor. In the case of a researcher, the foreseeable risk for a researcher could be avoided by applying an anonymity protocol that would excise the identity information.

Where circumstances of error are involved, if the researcher possesses any duty of care it is to follow the protocol (standards) for such noninvasive, non-interactive research, for example, to accurately and within its relevant context present the conversation or other activity or content so observed and recorded. Even if this duty could be established it would be judged by a standard of reasonableness, not under a no-fault or strict liability rule. Researchers have a scholarly duty to "get it right" but this does not necessarily equate to a standard of infallibility. Quantitative research recognizes the concept of error and deviation. Similar standards may apply in the context of such research when plaintiffs seek to claim that such deviation forms the basis of a compensable harm. Was deviation from perfection in the observation and collection of such a character so as to establish a duty and breach of that duty? A plaintiff would need to establish through the expert testimony of other researchers that the deviation was unreasonable. The public policy in favor of the progress of science may also work to prevent a court from assigning a duty of care in cases where the researcher acted within the parameters of the protocol where acceptable levels of error are allowed.

Public policy arguments may also drive the analysis where short of the obvious danger of the Gun-For-Hire scenario, the suspicious nature of the content included by the researcher is no less apparent. In Eimann v. Soldier of Fortune Magazine, Inc., the federal appellate court concluded that the following advertisement, unlike the "Gun-For-Hire" advertisements in Braun and Norwood, did not pose a foreseeable risk: "EX-MARINES-67-69 'Nam Vets, Ex-DI, weapons specialist-jungle warfare, pilot, M.E., high risk assignments, U.S. or overseas" (Eimann v. Soldier of Fortune Magazine, Inc., 1989, p. 832). The court discussed the social utility of the advertisement and the negative impact imposition of a duty of care would produce: "Given the pervasiveness of advertising in

our society and the important role it plays, we decline to impose on publishers the obligation to reject all ambiguous advertisements for products or services that might pose a threat of harm… The burden on a publisher to avoid liability from suits of this type is too great: he must reject all such advertisements. The range of foreseeable misuses of advertised products and services is as limitless as the forms and functions of the products themselves. Without a more specific indication of illegal intent than Hearn's ad or its context provided, we conclude that SOF [Soldier of Fortune] did not violate the required standard of conduct by publishing an ad that later played a role in criminal activity" (Eimann v. Soldier of Fortune Magazine, Inc., 1989, p. 838, emphasis original). It is logical to conclude that a court would see as much if not more value in the progress of science that the dissemination of proceeding paper, journal article, book chapter, publication represents. Concluding that researchers possess a foreseeable risk would garner a similar undesirable result: researchers would avoid including any suspicious, controversial, or otherwise unusual forum content in their findings, or forego inquiry into entire subject matters or genres of online speech. Such result is not in the public interest.

Assuming the rare instance where a duty is present, was the duty nonetheless breeched? If breach occurs, with what result does it occur? Is the breech the legal cause of the harm and what is the compensable harm? Short of defamation or invasions of privacy discussed or the clear and volatile circumstances of the crime victim or Gun-For-Hire cases in subsequent articles, what is the harm in being misquoted, for a researcher to claim that X's favorite movie is *Brokeback Mountain* not *Broken Arrow*, or *Mystic PizzaMystic River*, or subject Y's favorite sexual position or preferred frequency is confused with X's, or is different than the one reported by the researcher? Errors do occur, but did the researcher follow the proper error protocol while transcribing her notes or while subjecting the downloaded forum snippets to software based content analysis? Even if protocol was not followed, can the plaintiff forensically reconstruct the process to prove unreasonable deviation from the protocol? Moreover, the plaintiff subject would need to demonstrate a compensable injury occurred. (If an error occurs in the presence of negligence a false light claim is be possible. False light is a subset of the invasion of privacy tort and is discussed in a subsequent article.) The dollar value (compensable

harm) placed on these errors is likely small, or certainly smaller in relation to what it would take to retain counsel to litigate such disputes.

The reality of this factor in risk assessment cannot be overlooked as it is typical for personal injury lawyers to operate on a fixed percentage fee, where compensation is owed only if the claim is successful. Thus personal injury lawyers limit their personal docket to those cases that have a likelihood of success. Success equates to dollars, not just a winning verdict. The potential for recovery must be expected to exceed the outlay, the attorney's cost in time of pretrial, trial and post trail preparation and execution. The attorney's fee is typically between thirty to forty percent of the total award. Other costs must come from the plaintiff's share of recovery as well, that is, the cost of retaining expert witnesses to testify as to the duty of care (acceptable level of behaviour) and breach of that duty (deviation from the protocol by the researcher), costs of recording and transcribing a deposition, and so on. There must be a reasonable opportunity to recover all of these costs or the attorney is unlikely to undertake representation. As a result, the practical logistics of litigation serves to further minimize the risk of a legal proceeding even in the rare instance where some risk exists.

Finally, if damages are extant and determinable, a court would have to conclude that the breach of protocol was in fact the legal or proximate cause of the harm. This is a significant "if" as courts especially appellate courts often consider public policy impact that the imposition of liability would have against similarly situated plaintiffs and defendants in the future. In other words, if researchers could be sued for errors in phrasing or reporting mischaracterization (short of defamation, invasion of privacy, etc., see discussion below and in a future article, respectively), would it encourage the filing of unsound or frivolous claims by subjects or otherwise have a negative effect on the progress of research, that is, by chilling the desire of like-minded researchers fearing like result who then forego similar research projects in the future?

In other cases of measurable physical harm such as the online predator scenario, public policy considerations may also lead a court to conclude that the dissemination of postings by forum participants was not the proximate cause of harm that did result. Assessment of foreseeability is part of the proximate cause analysis as well. Moreover,

the concept of superseding cause may break the link of event and excuse the researcher for any liability for the harm resulting from the criminal act of another, in this example the criminal conduct of the predator.

So too, the case law of negligent publication demonstrates that liability is rare except where circumstances increase the magnitude and likelihood of harm, i.e., the criminal defendant became aware of otherwise unknown name of crime victim or Gun-For-Hire advertisements. In other cases of measurable physical harm such as the predator scenario described above, the public policy considerations may also lead a court to conclude that the dissemination of postings by forum participants was not the proximate cause of harm that did result. The link in terms of the law is too tenuous. Moreover, the concept of superseding cause may break the link of event and excuse the researcher for any liability for the harm resulting from the criminal act of another, in this example the criminal conduct of the predator.

The more likely remedy for "negligence" in the academy is some sort of censure by colleagues for the less than acceptable deviation from proper research protocol. Censure would take two forms, informally by fellow researchers through the peer review process or formally by disciplinary or other action for the violation of a policy regarding research integrity, Institutional Review Board practices, and/or responsible conduct of research. A court might likewise view either process as a more efficient method of regulation-response than pursuit of remedy through the courts.

Conclusion

In assessing the likelihood of success for harm based upon negligence, it is unlikely that a court would conclude that the duty of care owed by researchers to either the research subject or to third parties is complete accuracy. Attention to protocol regarding reporting errors would also eliminate risk where the defendant is defamed or otherwise harmed (negligence). A researcher would need to deviate in a demonstrable way from the standard protocol in reporting and that deviation must be proved to have caused measurable harm. Likewise a court would not conclude that the duty of care placed upon researchers applies to the range of all possible harms, however remote. Again the issue of damages would need to be proved. Moreover, the amount of the damages would need to

be of an amount that would make legal process worthwhile for an aggrieved plaintiff and his or her attorney. Finally, a court may refuse as a matter of law or public policy to hold that a researcher is responsible for harm that might result from the dissemination of forum content. The negative impact of turning all possible harms into foreseeable ones is obvious. Without this protection the risk-averse researcher would be reluctant to include such content resulting in a decrease in the rigor or robustness of inquiry.

The following guidelines can be used to assist the researcher in avoiding the possibility of claims made on the basis of negligent content. Follow standard methodology for error reduction to ensure that the duty of care is not breached; and, where danger of physical harm is possible but remote, avoid use of identifying information to ensure that a legal nexus does not exist between the breach of a legal duty and the resulting harm.

References

Black's Law Dictionary (8th ed. 2004). St. Paul, Minnesota: West Publishing Co.

Jansson, R. L. (2003). Researcher Liability for Negligence in Human Subject research: Informed

Consent and Researcher Malpractice Actions. *Washington Law Review*, 79:229-263.

Lee, B. L., *Modern Tort Law: Liability and Litigation* (2d ed. 2005). St. Paul, Minnesota: West Publishing Co.

Lipinski, T.A. (2006). Emerging Tort Issues in the Collection and Dissemination of Internet-Based Research Data, *Journal of Information Ethics*, Fall: 55-81.

Morreim, E. H. (2003). Medical Research Litigation and Malpractice Tort Doctrines: Courts on a Learning Curve. *Houston Journal of Health Law & Policy*, 4: 1-86.

Restatement (First) of Torts. (1934). St Paul, Minnesota: American Law Institute.

Restatement (Second) of Torts (1965). Philadelphia, Pennsylvania: American Law Institute.

Cases

Braun v. Soldier of Fortune Magazine, Inc., 757 F.Supp. 1325, aff'd 968 F.2d 1110 (11th Cir.1992), cert. denied 506 U.S. 1071 (1993).

Eimann v. Soldier of Fortune Magazine, Inc., 880 F.2d 830 (5th Cir., 1989), cert. denied 493 U.S.1024 (1990).

Guebard v. Jabaay, 42 N.E.2d 751 (Ill. App. Ct. 1983).

Hyde v. City of Columbia, 637 S.W.2d 251 69 (Mo. App., 1982), cert. denied 459 U.S. 1226 (1983).

James v. Meow Media, 300 F.3d 683 (6th Cir. 2002), cert. denied 123 S. Ct. 967 (2003).

Juarez v. Boy Scouts of America, Inc., 97 Cal. Rptr. 2d 12 (Cal. App. 2000).

Noah v. America Online, Inc., 261 F.Supp.2d 532 (E.D. Va. 2003), affirmed sub. nom. *Noah v. AOL-Time Warner, Inc.*, 2004 WL 602711 (4th Cir. Mar 24, 2004) (unpublished).

Norwood v. Soldier of Fortune Magazine, Inc., 651 F.Supp. 1397 (W.D.Ark., 1987).

Perna v. Pirozzi, 457 A.2d 431 (N.J. 1983).

Pugsley v. Privette, 263 S.E.2d 69 (Va. 1980).

Rowland v. Christian, 69 Cal.2d 108 (1968).

Sanders v. Acclaim Entertainment, Inc., 188 F.Supp. 2d 1264 (D. Colo. 2002).

Smith v. Linn, 563 A.2d 123 (Penn. Super. 1989).

Spenceley v. M.D. Anderson Cancer Center, 938 F. Supp. 398 (S.D. Tex. 1996).

Webb v. Desert Seed Co., 718 P.2d 1057 (Colo. 1986).

Creating a Web of Attribution in the Feminist Blogosphere

E. Hvizdak
School of Information Studies
University of Wisconsin-Milwaukee
PO Box 413
Milwaukee, WI 53202
ehvizdak@uwm.edu

Abstract

Through an examination of 143 women- and feminist-authored blogs and a review of the literature surrounding feminist perspectives of copyright, this paper seeks to demonstrate that a critical feminist perspective would be a valuable addition to the ongoing debates of copyright legislation. Current feminist perspectives primarily critique the law's privileging of the autonomous author-genius by citing women's historical participation in collaborative activities, deconstructing binaries based on that of creator/user, and citing the dialogic nature of cultural production. These theories might purport that feminists and women would emphasize information sharing over ownership as a response to formal copyright law. However, 38 % of the blogs expressed some type of copyright-protected status, indicating that creators in the digital realm seek to contribute to cultural production while at the same time retain authorial credit, made especially evident through the high rate of requests for attribution in future uses of the bloggers' works. A feminist perspective would be useful in helping future copyright legislation to negotiate these needs.

Introduction

Digital self-publishing is rapidly expanding to embrace authors from a variety of backgrounds, experiences, and identities, and especially those traditionally ignored, underserved by, or misrepresented in the mainstream media. Two such groups turning to digital media in search of expression and voice are women and feminists. As this electronic sphere gains in popularity, certain social norms are being created to regulate

behaviors and ensure that all are welcome to create, continuing intellectual progression. One such social norm involves copyright practices. As much publishing on the web relies heavily on the linking or quoting of others' creations or information, and because it is easier in the digital realm to copy and paste someone's work without the original creator's knowledge, norms are being created to ensure that works can be shared and that people can continue to create without fear that their works will be used unfairly. These norms are also created in reaction to formal copyright law. Such law protects works created by a single, autonomous author for economic gains, which are qualities many web publishers feel do not fit their creations. To ensure that voices will not be silenced when copyright law does expand to place additional formal restrictions upon digital works, it is essential that women's and feminists' practices and perspectives of authorial attribution, non-commercial creation, the breaking down of the boundary between author and user, and the close relationship of author and work are included in future legislative discussion.

Current copyright law in the United States embraces the idea that a single, autonomous author, free from influence or help by others, creates each potentially copyright-protected work. This concept emanates from the "heroic self-presentation of Romantic poets" (Woodmansee & Jaszi, 1994, p. 3) that developed some 200 years ago. While it is this romantic idea of autonomy that continues to inform legal arguments in favor of increasing copyright protections today, the "author" concept first came to the fore in the Statute of Anne of the 1700's (Jaszi, 1994, p. 32). This piece of English copyright legislation sought to protect capital gains of publishers and booksellers, not the writers themselves, in a world that was increasingly dominated by new printing technologies. Until 1750s Germany, the author "was still being represented as just one of the numerous craftsmen involved in the production of a book" (Woodmansee, 1994, p. 15), and only deemed an original creator when the economic status of those holding printing power were at stake. Edward Young's 1759 *Conjectures on Original Composition*, furthered by Wordsworth's *Essay, Supplementary to the Preface* of 1815, helped the concept of the Romantic author to emerge. It states, "Genius is the introduction of a new element into the intellectual universe: or, if that be not allowed, it is the application of powers in objects on which they had not before been exercised"

(Wordsworth, as quoted in Woodmansee, 2004, p. 16). The law parallels this notion that the author creates works in a sort of vacuum, without external influence.

Historical analyses of copyright legislation also closely examine the term "work" used in section 102 (a) of Title 17 U.S. Copyright Law (17 U.S.C.). Christopher May and Susan K. Sell paraphrase Peter Jaszi when they explain that the change in copyright law to this term of "work" implies an attachment of commodity to authorship, invoking the Lockean principle that "work should be rewarded with property" (May & Sell, 2006, p. 102. See also, Coombe, 1998, p. 219-220). In other words, mental "work" emanating from the individual, autonomous genius is to be acknowledged and privileged with protection leading to economic gains. This, of course, is the "intent" of the creation of current US copyright law. Article 1, section 8, clause 8 of the United States Constitution specifically states that copyright exists "To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." As long as creators are given monetary or economic rewards for their pieces of mental and/or physical work, made possible by granting temporary sole ownership over that work to the creator, then he or she will continue to create, thereby contributing to the progress of culture. This progress is insured by the eventual release of these works into the public domain, where they become available for others to borrow from in order to produce their own creations.

According to the U.S. Copyright Office's website, "Only the author or those deriving their rights through the author can rightfully claim copyright" (U.S. Copyright Office, 2006). Copyright is extended to any "original" creation in "fixed form," with or without registration (17 U.S.C.). "Fixed" implies that the sign is closed and final, free from outside influence and existing independently of any future change or adaptation. Works produced by multiple individuals are protected only under the "Works of Joint Authorship" section of United States Code Title 17 (17 U.S.C. 201). This principle, however, does not acknowledge a creation produced and shared equally and collaboratively amongst individuals. Instead, "the co-author must contribute original expression that could stand alone as copyrightable subject matter" (American Bar Association, 2006, p. 7). Each part of a "joint work" must contain separable, independent,

fixed signs, each created by a single individual, which still invokes the concept of one individual author-genius per one creation.

Copyright and Feminism?

Feminist legal theory has hotly criticized copyright law's privileging of the autonomous author, though little is written on the subject compared to other traditional "women's" legal issues, such as sexual and employment discrimination, pornography, family law, reproductive health, and so on (See for example, Taylor, Rush, and Munro, 1999). Citing the historical participation of women in collaborative activities, the dialogic nature of cultural production, and the power structures inherent in an owner/owned binary, the writings on feminism and copyright primarily critique the notion of the individual, autonomous author privileged in copyright law. Ann Bartow suggests that women are more likely to participate in collaborate activities, such as quilting, knitting, or cooking, activities that produce domestic "works" not generally protected by Title 17, section 102(a) (Bartow, 2007). Additionally, this collaborative effort is ignored by the law's protection and sole recognition of the single, autonomous author. Bartow goes on to explain that women who seek copyright protection violate the feminine social norms of caring, sharing, and nurturing, therefore deterring women from seeking this protection (Bartow, 2007, p. 33). Craig furthers this when she states, "Employing the notions of dialogism and the relational self that have emerged from feminist scholarship…we can reimagine the author not as source, origin, or authority, but rather as participant and citizen" (2006, p. 234). But, as described, existing copyright law privileges and protects works that emanate from the single, autonomous individual, extending no thought, acknowledgement, or protection to the feminist notion of collaborative, relational or dialogic cultural production.

Feminist theory also deconstructs the binaries present in copyright doctrine, exposing patriarchal power structures. Dan L. Burk cites dualisms such as mind/body and nature/culture, (Burk, 2006) while Craig cites laborer/free-rider, creation/reproduction, and author/user (Craig, 2006). Each of these binaries holds the characteristic of one side being privileged over another, or one side being "inferior and feminized" (Burk, 2006, p. 11). For example, the most prevalent binary, author/user, is invoked to determine

infringement. The author is the creator, the sole owner of the work, and that who has control; the user, in any attempt to become involved with the piece, such as changing or borrowing from it, becomes an infringer and is punished by law. The user must separate him or herself from the author and his or her work, becoming an outside spectator rather than an active participant. Not only is this binary problematized by the assertion that culture, and therefore creation, works in a dialogic manner, but also in the fact that it is the public, the audience, or the user that makes a work economically viable or worthy of copyright protection (Zemer, 2007, p. 5-6). In other words, without the user or consumer, the author or creator would have no reason to call him or herself a unique, autonomous, author-genius under copyright protection.

Additionally, the feminization and "othering" of one side of the binary parallels the claiming of copyright, patents, or ownership of non-Western and indigenous creations and knowledge. Written culture is copyright-protected, while oral culture is not; individual, original works are copyright-protected, while collaborative, culturally produced works are not. Additionally, Kembrew McLeod uses the example of "world music," stating that because these beats and songs are not "original" works by one individual, they are considered free reign under law (McLeod, 2001, p. 243-244). Those that have the education, writing and reading ability, and knowledge of copyright law are then able to use and therefore protect, for monetary gains, the uses of these beats in their own creations. The presence of the author/user binary, if Burk and McLeod are correct in their assertions that the latter side of the dichotomy is feminized or "otherized," leaves those that lack the access to knowledge of U.S. and international copyright law at a distinct economic disadvantage.

Some authors claim that a true feminist future for intellectual property law would be to greatly lessen its protections or to get rid of it all together. Bartow explains that "low protectionism" would benefit women as "authors, intermediaries, and consumers" (2007, p. 24). Women authors and creators could seek royalties and protection for their work without being seen as selfish or greedy; women intermediaries would be allowed to bring works together to "advance the cause of sisterhood;" and women consumers would benefit from having more works to choose from to borrow, reproduce, and distribute without fear of punishment. Deborah Halbert takes this a step further when she states that

"A feminist future would eliminate the law of intellectual property, which is too often used to halt creativity, and replace it with an understanding of the community in which one creates" (1999, p. 119). Thus, these theories imply that copyright law should move away from the privileging of the autonomous author-genius and towards an acknowledgement of the collaborative and relational nature of work creation and information exchange.

In The Blogosphere?

The weblog is a collaborative form of media that emphasizes dialogue and deconstructs the dualisms and binaries present in copyright law through prevalent sharing, quoting, and linking of information. Weblogs, or "blogs" as they are more commonly called, are emerging rapidly as a site of creation by women and feminists, among many others. Statistics from the UK's Guardian Unlimited include: "A recent estimate put the number of feminist blogs at 240,000, but, given that this posited the number of "active" worldwide blogs at 4m (some figures put it as high as 27.2m), and the proportion of women who are self-described feminists at 10% (a British survey this month produced a figure of 29%) the true figure could be much higher" (Cochrane, 2006. See also, Tobias, 2005). Many of these women and feminists are turning to the web based on the fact that their voices are being ignored or misrepresented in mainstream media. For example, the organizers the October 1997 Million Woman March, a gathering of primarily African-American women and men, turned to the Internet to spread information about their event (Everett, 2004). The organizers subverted the digital divide by relying on women that did have access to the Internet to print out and distribute information to those that did not, demonstrating the power of the Internet to mobilize women both onscreen and off. Virtually ignored by mainstream mass media, the event planners thrived in the Internet sphere through websites and weblogs, drawing hundreds of thousands of people and generating 21.7 million dollars for the city of Philadelphia. Bartow (2005) also explains that women, such as new and stay-at-home mothers, use the Internet to connect with one another in a relatively anonymous sphere. Needing a space of their own outside of mainstream media, where connections and relationships can be formed quickly and

cheaply across geographical and social boundaries, many women and feminists have turned to the blogosphere.

Blogging and self-publishing on the web can be used to illustrate feminist reaction to copyright (See also: Bartow, 2006). To start, the blog's hypertextual nature highlights the relational nature of information and cultural production (Burk, 2005). Meaning is not drawn solely from the article or posting itself, but is often gathered from external sources made evident through "weblinks." The ability of the reader to link separate pieces of already-existing information together in a sequence of his or her own choosing problematizes the boundaries of a "creation" and also the binary of creator and reader. In this sense, the reader becomes the creator, reminiscent of Barthes's "writerly" texts (Barthes, 1977). According to Sonia K. Katyal, the reader also becomes the writer in the creation of online slash fan fiction, texts that rework mainstream, copyright-protected published works (Katyal, 2006). Specifically, these texts re-imagine the heteronormative, leading male of the story as being involved in homoerotic relationships with other characters (the term "slash" coming from the "m/m" relationship). The high number of female authors appropriate and deconstruct male dominance, placing the characters in a relationship where the power balance is equal (male/male) rather than in the original text, where power is unequal (male/female). This action reworks the gender roles made rigid by the copyright-protected original works in an attempt to show the fluidity of gender, and therefore the fluidity between reader and creator (Katyal, 2006, pp. 485-486). In each of these cases - the event website, online slash fan fiction, and especially blogs - the boundaries between reader and author and even the boundaries surrounding each "autonomous" author are broken down through the reworking of texts and the heavy exchange and redistribution of information.

Feminist theory might purport that women and feminist bloggers would not seek copyright protection for their works. Blogs are highly collaborative efforts, relying on information from external sources (news media, other individuals) to create meaning, and encouraging readers to add to the creation by posting comments. This is in stark contrast to the concept of the autonomous author as sole creator in copyright law, so bloggers might not see their blogs as created by a single person and therefore worthy of or needing protection. Additionally, rejecting copyright protection might align itself with feminist

activity, subverting hierarchical patriarchal power by emphasizing and encouraging collaborative creation. The statistics on this subject are very telling regarding the ways that women and feminists view and practice copyright in the blogosphere, and these data can then be used to inform copyright legislation in general.

Methodology and Research Questions

This research focused on samples from two feminist "blogrolls," noting there are many from which to choose. The first was from "Feminist Blogs" (Feminist Blogs, 2007). Each of the 53 blogs listed was examined, excluding nine that were unavailable and therefore not counted in the results. The other sample comes from the "Feminist Blogs US" blogroll listed on the blog "Laurelin in the Rain" (Laurelin in the Rain, 2007). A total of 34 blogs were listed. Two were unavailable and five were duplicated from the first blogroll, and were therefore not included in the results. The two blogrolls together produced a sample total of 71 blogs. For my second set I sampled blogs listed in the "Women Bloggers Webring" (Women Bloggers Webring, 2007). Sampling from this site gave an idea of the copyright practices of women that blog as "women" but not specifically feminists. From the 295 blogs listed, I examined every fourth blog, beginning with the first listed. All were available, and none were duplicates of the feminist blogs, which gave a sample of 72 blogs. Notably, IRB approval was not sought, after consultation with the board administrator, who excluded blogs from the realm of reviewable human subjects research.

The following questions were considered when surveying the blogs: What types of copyright protections do women and feminists seek? Is a copyright protection symbol/statement or link to said symbol/statement present on the initial homepage? If yes, what type of copyright protection? If the blog is licensed under Creative Commons (or another similar license), what type, and what are the aspects of each license? If a self-authored copyright statement is present, what are the main elements or protections?

Results and Discussion

Of 143 blogs surveyed, 55 had some type of copyright statement or a link to it present on the homepage, while 88 did not. These data can be further broken down into women's and

feminists' blogs. Women-authored blogs expressed copyright-protected status in 31 out of 72 instances, while feminist blogs expressed it in 24 of 71 instances.

The types of copyright-protection notification varied greatly over the blogs. Simply using a copyright symbol (©) with a statement of the creator's name, blog's name, and/or year was the most popular form of copyright-protected notification in each of the two sets of blogs. Just a few examples include:

©2005- 2007 LIFE WITH HEATHENS AND JCB - ALL RIGHTS RESERVED. ANY UNAUTHORISED COPYING, REPUBLISHING, OR USE WILL INFRINGE COPYRIGHT" (Life With Heathens, 2007)

ALL CONTENT © 2004, 2005, 2006 IT'S MINE! ALL MINE! (Suburban Lesbian, 2007)

(Copyright) © 1992-2007 The Insane One (aka Teli). Please don't steal my stuff. ☐Validate my XHTML // CSS at your own risk. Other disclaimers coming soon... (Mild Insanity, 2004).

A Creative Commons license, explained below, was the second-most popular form of notification in each blog set, while a copyright symbol and a further, self-authored statement was the third most popular in each blog set. In five instances in the feminist blogs, a Creative Commons license was combined with a copyright symbol; this occurred only once in the women-authored blogs. The women-authored blogs also included one instance of Creative Commons combined with Copyscape, and also one instance of Copyscape by itself (Copyscape, 2007).

Creative Commons deserves further explanation, as it is increasingly defining copyright practices in the digital realm. This licensing website allows a creator to choose one of six "Some Rights Reserved" licenses comprised of any combination of four restrictions: Attribution, Noncommercial, No Derivatives, and Share Alike (Creative Commons, 2007). A banner is situated in a visible place on the license-holder's webpage, and clicking on it will bring the reader/user to the actual terms of the license, both in "everyday" speak and also in legal terms. A June 2006 estimate put the number of

Creative Commons licensed webpages at 140 million - but of the 12 billion sites on the web, this accounts for only 1.2% (Lastowka, 2007, pp. 80-81). This percentage, however, is far less than that of the women's and feminist blogs examined. Creative Commons licenses were present in 22 out of 55 blogs total (10 of the women's blogs and 12 of the feminists' blogs), or 40% of the time, demonstrating that women and feminists use Creative Commons on a far more regular basis than the general web population. Lawrence Lessig, one of the founders of Creative Commons, makes clear in his text *Free Culture* that these licenses are not built in an effort to work against existing copyright law, but instead to "make it easier for authors and creators to exercise their rights more flexibly and cheaply" (Lessig, 2004, p. 286). The idea is to allow works to be used and distributed by others as long as the original author's wishes are respected. This circulates more works in the creative sphere, relying on a level of trust between creator and user that all conditions are met.

In the survey of the distribution of specific chosen restrictions found in both the blogs' Creative Commons licenses and self-authored copyright statements, each restriction was counted once, and as most of the licenses and self-authored statements have more than one restriction, some blogs were counted more than once. The first, Attribution, is included as a default in every Creative Commons license and so was only counted when it occurred on its own. This feature went into effect in 2004 when it was discovered that 97% of Creative Commons license holders had included it as one of their chosen restrictions (Lastowka, 2007, p. 80). Attribution also occurs at a high rate in the self-authored statements of both women's and feminists' blogs: 6 out of the 9 blogs, or 67% of the time. It was trumped only by "Permission" (a user must gain permission from a creator before reuse), which occurred in 7 of the 9 blogs, or at a rate of 78%. Permission is also built into all Creative Commons licenses, and so was not counted as a separate factor in these instances.

The high rate of requests for attribution and permission (both in the self-authored statements and as they are built into the Creative Commons licenses) seems very much in line with feminist action: to foster relationships, display connectivity in creative endeavours, and dismantle hierarchies of knowledge-holding by citing where a piece of information came from. Requesting that one's work is cited in subsequent creations can

also be seen as a way of building or gaining social capital in what Lastowka calls a "social reputation market" (Lastowka, 2007, p. 43). Rather than seeking monetary gains for their works, creators instead seek social gains. The more their name is out there, the more the readers, users and general audience will include them in social groups and networks, ask them to produce more information, and in general provide them with a wider social circle (whether that be in person or over the Internet), rather than providing monetary gains. Again, this would seem to embrace relationships and social mindsets - creators are looking for an increase in their interactions with others and to hold a place in this web of social and creative production. Additionally, it quite explicitly breaks down the binary of author/user, emphasizing that the work was produced with the help of external forces and will continue to inform more works in the future.

This preference of social over monetary gain is emphasized by the high rate of "Non-commercial" usage restrictions. This occurred in 4 of the 9 self-authored statements, or 44% of the time, and it was listed as a restriction in 16 of 22 Creative Commons licenses, or 73% of the time. The non-commercial restriction states that a person may print out or use a creator's work, but may not sell or profit from it. This, combined with the high use of the attribution restriction, indicates that creators are trying to protect their potential social gains, not their economic gains. It could also indicate the creator's wishes to contribute only intellectually, and not capitalistically, to the intellectual growth of society. This non-commercial restriction could also be seen in terms of the cultural shift to create works via the internet. This type of information creation and distribution does not require a publisher, instead only requiring a person with access, skills, and the desire to speak. Capitalistic gains for a publisher do not even come into play in this sphere. This, combined with the fact that access to most of the content on the web is free, could be seen as the creators' reclaiming of copyright protection from publishers.

However, bloggers made no indication that they themselves would not use their works in an economic manner. For example, excerpts from Darkdaughta's copyright statement are as follows (Darkdaughta, 2005):

our worlds converge more often than you think. so, unless you want to see me staring at you from a seat in the audience at a conference/panel/symposium you've organized to discuss power, oppression, co-optation, legitimacy, knowledge or some such other drivel, as you spit my words back at me, do not selfishly mine my intellectual and creative resources for your own economic well-being and stability.

same fyi for paid community organizers who are working against oppressive odds in professional environments where the sort of work i do, facilitation/education/dissemination of thought through spoken and written word, is becoming de-skilled, whittled down to under-paid work divested of any purpose beyond seeing to the programming needs of middle[activist]management who report to academic, government and foundational funding sources.

…folks were translating info, ideas, energies this cash poor, working class woman with intellect was offering into cash making opportunities for themselves inside community and they weren't at all interested in sharing what they were cookin' up with me!

Similarly, Bonni at bonni.net states on her blog (Hall, 2007):

Content and design copyright © Bonni Hall, 2000-2007. All rights reserved. No unauthorized use or reproduction is permitted. It's not that I don't love you. It's just that I don't want you to plagiarize my original writings or take my graphics or layout, which I worked very hard to produce. Thanks for understanding.

If you like my graphics, allow me to say that I am, among other things, a professional web graphic designer and would be more than happy to make a website template or graphics set for you of your very own, and I'm pretty darned affordable.

Each of these women make explicit that they worked hard on their content, and they do not want anyone else to receive monetary gains based on this work. However, this does not mean that the women are opposed to making money from their work. Bonni is a

graphic designer who posts material on the web - she still embraces the monetary gains of copyright while sharing the looks, but not use, of these graphics with others. Darkdaughta is opposed to the forces ("academic, governmental, foundational funding") that get in the way of paying creators what they are due (Darkdaughta, 2005). Interestingly, Bonni Hall at bonni.net seems to conform to Bartow's theory that women that seek protection might feel they are breaking the feminine caring and nurturing social norms by claiming ownership. Bonni states, "It's not that I don't love you" and "Thanks for understanding," as though she is apologizing for protecting her works (Hall, 2007). In either case, the women do not make explicit that they are opposed to monetary gains for themselves, but are instead opposed to others using their works to do so. The women lose out on social credit as well as monetary credit when others use their works for commercial purposes.

Interestingly, though, of the 22 Creative Commons-licensed blogs, only a one license-holder had an "Attribution Only" license. This is very telling in terms of women's and feminists' views regarding ownership and use of others' works. As all of the other blogs besides one had further restrictions, it demonstrates that these creators not only want to be acknowledged for their works, but also want to exercise control over the present and future uses of their creations. After the single "Attribution Only" license, Share Alike was the least common restriction found among Creative Commons license holders, held by nine of the 22 license-holders. This is distributed fairly evenly between women's and feminists' blogs, with 5 occurring in the women's blogs and 4 in the feminist blogs. Share Alike requires individuals using licensed content to then license those derivative works under the same restrictions, meant to ensure the continuation of information sharing and the creative furthering of culture. This is similar in the Attribution sense, as it ensures that the connections between pieces of information are made known from derivation to derivation, creating a web.

"No Derivatives" was the second most-common chosen restriction, occurring in 10 of 22 cases. What is so interesting about this number is the uneven distribution between women's and feminists' blogs. This restriction occurred 7 times in feminist blogs, and only 3 in women's. The restriction in itself is the most interesting of the four as well, as it calls for the most protection. Users are free to read and distribute this information all that they want, but they are not allowed to make any sort of derivation of

that work, even if attribution is given. This, like current copyright law, keeps the creation a fixed sign, belonging to a single author. What is even more interesting is that this comes up more often in the feminist blogs. Halbert (1999), for example, seems to support the act of appropriation as a feminist way of resistance against rigid copyright laws. To allow others to create works based on your own would be to explicitly break down the binaries of reader/creator, allowing information to transform itself and evolve as each next person consumes it. Perhaps the feminist "No Derivatives" ratio is higher because these blogs involve a more activist tone, or are created with a strict purpose in mind to convey a very specific message that the creators do not want altered or misrepresented. Additionally, as feminism has not always been well received in any society around the world, perhaps these feminists are trying to protect their works from being parodied, slandered, or the like. If either of these is true, then the social reputation factor comes into play once again. No Derivatives is in place to protect the integrity of these politically charged and sometimes emotionally sensitive messages, thereby protecting the social well-being of the creator.

Feminist legal theorists of intellectual property and copyright law might claim that women would not want to assert copyright protections over their works, that they would want to emphasize and participate in collaborative efforts, and that their practices do not fit well with the patriarchally-structured binaries inherent in copyright doctrine. 38% of the blog sample, however, would argue to the contrary. These women seek protection for the integrity of their works, and they want to be able to claim these creations as their own. They include their name on the blog with a copyright symbol; they license protections under Creative Commons; or they author their own restrictions regarding uses of their works. While 38% is not quite half of the sample, it is still a substantial number worth heeding. What should be emphasized in these women's and feminists' practices is the prominence on the blog of the name of the copyright holder (for example, the © plus the person's or blog's name) and more importantly, the high rates of attribution and noncommercial restrictions placed upon present and future uses of these blogs.

Copyright law does not privilege, protect, or recognize these elements. First, the changes in the 1976 revision made all works in tangible, fixed form protected, with or

without the presence of a copyright symbol, and one does not have to formally register their work with the copyright office for it to be protected by law (17 USC). 38% of women and feminist bloggers, however, decided to affix some form of notice of copyright-protected status, indicating that these bloggers do in fact desire a claim to individual authorship. The relative anonymity that can occur in the digital realm and the lack of need to affix a copyright symbol or authorial information to a work can make it impossible to attribute a work in the first place. By affixing a copyright symbol and statement, these authors are exercising their attribution rights. In addition, they are preventing "orphan works," or those copyrighted works that hold little economic value but are still protected, and whose authors cannot be found to gain permission to release these into the public domain for creative use by others (U.S. Copyright Office, 2007, "Orphan Works"). To affix a statement of authorship to a work is, on one hand claiming ownership over the work, but also, if indirectly, helping to avert the orphan works crisis by making it easier to track down authors to ensure future usage of these works.

Copyright law does not fully protect the right of attribution, or what Roberta Rosenthall Kwall refers to as a "moral right," explaining that copyright law protects only "pecuniary" and not "personal" interests of the author-creator (2002, p. 989). Attribution rights are only found in VARA (Visual Artists Rights Act of 1990), DMCA (Digital Millennium Copyright Act of 1998), and section 43(a) of the Lanham Act (Lanham Act, 15 U.S.C. 22). VARA prevents "modifications and misattributions of certain works of visual art" (Kwall, 2002, p. 987). DMCA contains "anti-circumvention provisions to protect the integrity of the digital protections that enclose and encode those works" - in other words, rights metadata cannot legally be removed or bypassed for use and distribution (Lastowka, 2007, p. 45). The Lanham Act prevents "false designation of origin" (Kwall, 2002, p. 988). However, VARA only protects attribution rights in certain works of visual art, not all creations (Kwall, 2002, p. 994); even with the DMCA piracy and illegal downloading runs rampant (Lastowka, 2007, p. 46); and it is often difficult to determine if the "false designation of origin" cited in the Lanham Act was an intentional act or if consumers were simply confused between brands (Kwall, 2002, p. 1026). Addtionally, section 106 of U.S. Title 17, that which protects the rights of authors, makes no mention of a right to demand attribution (Kwall, 2002, p. 996), and Fair Use section

107 does not have a requirement to attribute in the works that are used and distributed (Lastowka, 2007, p. 84).

Copyright infringement liability still occurs when the author of a plagiarized work is revealed - the infringement lies not in lack of attribution, but in the actual act of plagiarizing (Kwall, 2002, p. 998). The law makes little, if any attempt to protect the personal reputation rights of an author, instead punishing those who simply participate in the act of plagiarizing, whether attribution was given or not. This relates to the high rate of chosen non-commercial restrictions in women's and feminists' blogs. One can still make money copying and distributing the works of others, which comes back to copyright law's original existence to protect the economic gains of the publisher, not the personal gains of the author. Creative Commons licenses and self-authored statements remove the power of hierarchical publishing companies to claim distribution rights and economic ownership of a work, therefore returning all rights back to the creator. This restriction also adds an additional safeguard to one's works so that those works will not be used improperly in the economic sphere.

Self-publishing in the digital realm is on the rise, and takes the form of blogs, fan fiction, personal web sites, posting one's academic articles, and the like. Feminists, women, and others marginalized by mainstream media that have access to the Internet are especially turning to the web as a means of support, expression, and mobilization. Within this sphere, copyright norms are being created and exercised regarding how one protects and allows for the use of their creations. Specifically, those who provide a statement of copyright protection or licensing on their page provide the name of the author or blog and/or request that those who use the work provide attribution on any derivations or distributions. These actions signify a return of copyrights from the publisher to the author, forging a stronger connection between work and author and therefore facilitating stronger relationships between reader and creator. The requirement of attribution and the embracing of copyrights by women in the digital sphere can be seen as an attempt to strike a balance between protection of the integrity of one's work and also the desire to contribute to the progression of cultural knowledge. The call for attribution is forming a web of information production, demonstrating the connections between each piece of

produced work while providing for social gains and reputation in the form of authorial naming.

Conclusions

Feminists and women's rights activists have struggled for the rights of women and others outside of mainstream conceptions of gender for a number of years. To reject notions of ownership of a work would also be to deny agency to these women and feminists - agency that activists have fought to bring to the forefront for generations. Attribution is a way of bringing these two sides of the copyright binary together - it allows one to retain control over his or her creation and therefore obtain social gains while at the same time emphasizing the collaborative nature of knowledge production and the forging of social relationships. While many of the authors of texts on feminist perspectives of copyright call for a change in the law to embrace traditional "women's" collaborative works such as quilting or cooking, the feminist focus should instead work to negotiate the author/user binary so that shared knowledge production is encouraged and the rights of authorial ownership and attribution are ensured.

Creation in the blogosphere depends heavily on the linking and quoting of others' information and the subsequent trust that proper attribution will be given. As copyright law continues to expand to place more and more restrictions upon the uses of copyrighted works, authors of creations not falling under the category of a legally protected work - those that rely on collaboration, sharing, and a strong relationship between creator and reader to exist - threaten to be silenced. While it is true that creators will still want to profit from their works, it is essential, in this digital world lacking hierarchical publishing companies, that copyright law extend to provide moral or personal rights to authors in addition to economic rights. A feminist perspective would be essential in helping this legislation come to fruition. Feminist- and women-authored blogs provide a look into just a small segment of the intellectual property systems being developed outside of and in response to formal copyright law. Future research should aim to uncover additional examples of this type of information exchange so that they may be used in the development of a more creator- and user-balanced legal system.

References

American Bar Association. (2006). *What is a copyright*? Chicago: American Bar Association.

Barthes, R. (1977). Death of an author. In *Image, music, text* (pp. 142-154). New York: Hill and Wang.

Bartow, A. (2006). Copyright norms of the blogosphere: Linking, quoting, copying, and self-help. Paper presented at the 6th Annual Intellectual Property Scholars Conference, Berkeley CA. Aug 10, 2006. Available at: http://www.law.berkeley.edu/institutes/bclt/ipsc/papers2/Bartow.doc.

Bartow, A. (2007). Fair use and the fairer sex: Gender, feminism, and copyright law. *American University Journal of Gender, Social Policy and the Law, 14*(3), 551-584.

Bartow, A. (2005). Women in the web of secondary copyright liability and Internet filtering. *Northern Kentucky Law Review*, *Symposium Issue*, *32*(3), 1-59.

Burk, D. L. (2005). Feminism and copyright in digital media. In R. A. Spinello & H. T. Tavani (Eds.), *Intellectual property rights in a networked world: Theory and practice* (pp. 161-181). Hershey PA: Idea Group Publishing.

Burk, D. L. (2006). Feminism and dualism in intellectual property. *American University Journal of Gender, Social Policy and the Law, 15*(2), 183-206.

Cochrane, K. (2006, March 31). The third wave: At a computer near you. *Guardian Unlimited.* Retrieved August 1, 2007, from http://www.guardian.co.uk/g2/story/0,,1743520,00.html.

Coombe, R. (1998). *The cultural life of intellectual properties*. Durham, NC: Duke University Press.

Copyscape. (2007). Retrieved August 1, 2007, from http://www.copyscape.com.

Craig, C. J. (2006). Reconstructing the author-self: Some feminist lessons for copyright law. *American University Journal of Gender and Social Policy and Law, 15*(2), 207-268.

Creative Commons. (2007). *About: Choosing a license*. Retrieved July 16, 2007, from http://creativecommons.org/about/licenses.

Darkdaughta. (2005, March 10). *Copyright Stuff*. Retrieved July 14, 2007, from http://darkdaughta.com/copyright.html.

Digital Millennium Copyright Act. Pub. L. 105-304, 112 Stat. 3860 (1998).

Elkin-Koren, N. (2006). Exploring Creative Commons: A skeptical view of a worthy pursuit. In L. Guibault & P. B. Hugenholtz (Eds.), *The future of the public domain: Identifying commons in information law* (pp. 325-345). The Netherlands: Kluwer Law International.

Everett, A. (2004). On cyberfeminism and cyberwomanism: High-tech mediations of feminism's discontents. *Signs: Journal of Women in Culture and Society, 27*(1): 1278-1286.

Feminist Blogs. (2007). Retrieved July 14, 2007, from http://feministblogs.org.

Halbert, D. (1999). Poaching and plagiarizing: Property, plagiarism and feminist futures. In L. Buranen & A. M. Roy (Eds.), *Perspectives on plagiarism and intellectual property in a postmodern world* (pp. 111-120). Albany: State University of New York Press.

Hall, B. (2007). Copyright & legal information. Retrieved July 14, 2007, from http://bonni.net/copyright.html.

Jaszi, P. (1994). On the author effect: Contemporary copyright and collective creativity. In M. Woodmansee & P. Jaszi (Eds.), *The construction of authorship: Textual appropriation in law and literature* (pp. 29-56). Durham, NC: Duke University Press.

Katyal, S. K. (2006). Performance, property, and the slashing of gender in fan fiction. *American University Journal of Gender, Social Policy and the Law*, *14*(3), 463-518.

Kwall, R. R. (2002). The attribution right in the United States: Caught in the crossfire between copyright and section 43(a). *Washington Law Review*, *77*(4), 985-1032.

Lastowka, G. (2007). Digital attribution: Copyright and the right to credit. *Boston University Law Review, 87*(1), 41-90.

Laurelin In The Rain. (2007). Retrieved July 20, 2007, from http://laurelin.wordpress.com.

Lessig, L. (2004). *Free culture: How big media uses technology and the law to lock down culture and control creativity.* New York: Penguin Press.

Life With Heathens. (2007). Retrieved July 20, 2007, from http://www.lifewithheathens.blogspot.com.

May, C. & Sell, S. K. (2006). *Intellectual property rights: A critical history*. Boulder, CO: Lynne Rienner Publishers.

McLeod, K. (2001). *Owning culture: Authorship, ownership, & intellectual property law*. New York: Peter Lang.

Mild Insanity. (2007). *Copyright info & disclaimer*. Retrieved July 20, 2007, from http://www.mildinsanity.com/archives/2004/09/07/copyright-info-disclaimer/.

Suburban Lesbian. (2007). Retrieved July 20, 2007, from http://lifeinthesuburbs.blogspot.com/.

Taylor, B., Rush, S., & Munro, R. J. (Eds.). (1999). *Feminist jurisprudence, women and the l aw: Critical essays, research agenda, and bibliography*. Buffalo, N.Y.: William S. Hein & Co., Inc.

Tobias, V. (2005). Blog this! An introduction to blogs, blogging, and the feminist blogosphere. *Feminist Collections*, *26*(2/3), 11-17.

Trademark Act of 1946 (Lanham Act). P.L. 79-489, 60 Stat. 427. 15 U.S.C. Sections
        1051 et seq. (1946).

The Patent Copyright Clause, Constitution of the United States of America. U.S.C.S.
        Const. Art. 1 Sec. 8. Retrieved July 25, 2007, from
        http://www.law.cornell.edu/constitution/constitution.articlei.html.

U.S. Copyright Office. (2003). *Copyright Law of the United States of America and
        Related Laws Contained in Title 17 of the United States Code. Circular 92.* 17
        U.S.C. Retrieved July 21, 2007 from http://www.copyright.gov/title17/.

U.S Copyright Office. (2007). *Orphan Works*. Retrieved August 8, 2007, from
        http://www.copyright.gov/orphan.

U.S. Copyright Office. (2006). *U.S. Copyright Basics*. Retrieved July 21, 2007, from
        http://www.copyright.gov/circs/circ1.html

Visual Artists Rights Act of 1990. 17 U.S.C. § 106A (1990).

Women Bloggers Webring. (2007). Retrieved July 19, 2007, from
        http://l.webring.com/hub?ring=womenbloggers

Woodmansee, M. (1994). On the author effect: Recovering collectivity. In M.
        Woodmansee & P. Jaszi (Eds.), *The construction of authorship: Textual
        appropriation in law and literature* (pp. 15-28). Durham, NC: Duke University
        Press.

Woodmansee, M. & Jaszi, P. (1994). Introduction. In M. Woodmansee & P. Jaszi
        (Eds.), *The construction of authorship: Textual appropriation in law and
        literature* (pp. 1-13).   Durham, NC: Duke University Press.

Zemer, L. (2007). *The idea of authorship in copyright*. Burlington, VT: Ashgate
        Publishing Company.