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.