Exploring Ethical Issues in Data Mining: The Role of Collective Privacy

by

Kenna Jill Miskelly

B.Sc., University of Victoria, 2003

A Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF ARTS

in the Department of Philosophy

Exploring Ethical Issues in Data Mining: The Role of Collective Privacy

by

Kenna Jill Miskelly

B.Sc., University of Victoria, 2003

Supervisory Committee

Dr. Marsha Hanen, (Department of Philosophy)

Supervisor

Dr. Scott Woodcock (Department of Philosophy)

Co-Supervisor

Dr. Conrad Brunk (Department of Philosophy)

Departmental Member

Dr. Colin Bennett (Department of Political Science)

External Examiner

Supervisory Committee

Dr. Marsha Hanen, (Department of Philosophy)
Supervisor

Dr. Scott Woodcock (Department of Philosophy)
Co-Supervisor

Dr. Conrad Brunk (Department of Philosophy)
Departmental Member

Dr. Colin Bennett (Department of Political Science)
External Examiner

ABSTRACT

Data mining and other information technologies cause ethical concerns when they are used to categorize and discriminate. Even though there is an intuitive connection between privacy and personal information, it is hard to conceptualize the troubling issues raised by certain data mining applications in terms of privacy. This is largely due to the emphasis that traditional privacy definitions place on the value and protection that privacy provides to the individual. A notion of "collective privacy" emphasizes the broader social importance of privacy and provides philosophical clarity to the privacy issues raised by data mining. The policy suggestions that result from acknowledging the social value of privacy could benefit many in our society and work to fortify our privacy in this information age.

Table of Contents

Acknowledgements

To both of my supervisors, I would like to express my sincere appreciation. I thank Marsha Hanen for her guidance, advice, and patient, outstanding supervision. I thank Scott Woodcock for his insight, critical eye, and quick wit. I would also like to thank Colin Bennett for his critical assessment and his valued approval of this work. I thank Conrad Brunk for taking the time to read and respond to this thesis.

My success is due in no small part to the love of my family and friends. I thank my parents, the different members of my families, and my friends for their constant support. My husband Kelly, who is always ready to assist and encourage me, deserves my deepest gratitude. As does my son Vrai, who helped me in so many different ways.

Introduction

Our society has become increasingly networked with advanced communications and information sharing across diverse sectors. Personal data is now collected, stored, linked and aggregated as never before. This has led many to question whether or not new technologies have led to a loss of privacy.

Data mining is one such technology that is now ubiquitous, and although it can appear benign, a closer examination reveals complicated underlying ethical concerns. Within the business sector, this technology enables corporations to sort consumers into various profiles which can lead to de-individualization. Consumers who match the profiles deemed to be less profitable to a business may be denied incentives or services, or may be pressured to quietly end business with the organization. In the short term, this may lead to negative consequences for certain consumers. In the long term, this practice may cause serious societal harm if the profiles pervade our social consciousness and promote discrimination and de-valuing of certain groups of people.

This thesis questions the role that privacy plays in the ethical analysis of data mining. Data mining is a complex topic as it relies on largely public, non-intimate and anonymous information that is used within the business sector – a sector with its own goals and ends. Sifting through the ethical concerns of data mining can be complicated and authors disagree among themselves what the underlying issues are and how they should be addressed.

To help clarify this ethical uncertainty, I focus on the often overlooked intuitive and pragmatic connection between the social value of privacy and information technologies. I argue that traditional conceptions of privacy focus rather narrowly on the

importance of privacy to the individual. This ignores the intuitive connection between privacy, personal information, and the social costs often associated with information technologies. A shift to acknowledge the collective nature of privacy better situates the sentiments we have about the social value of privacy. This leads to a recognition of the greater social importance of personal information which, in turn, helps to clarify the ethical issues of data mining. Translating an emphasis on collective privacy into law and policy could also help address the influences and duties that businesses and organizations have in shaping the social nature of privacy.

I draw on a critique from the margins to provide support for the claim that privacy has important social value that is often ignored. Valuable social commentary is presented in an analysis of the impact that dominant privacy conceptions have had on Aboriginal self-determination. There is evidence that our laws and policies do not easily accommodate the social importance of privacy that is apparent in many Aboriginal cultures. A lack of attention to collective notions of privacy may contribute to the de-valuing of certain groups of people. The greater social importance of privacy may also be apparent in much of the way we commonly speak about information technologies. These considerations provide evidence that our laws and policies on privacy may be too narrow. Instead of focusing primarily on the relationship between privacy and the individual, there may be reasons to acknowledge our common sentiments regarding the larger social value of privacy.

Building on this social commentary, I will demonstrate how recognizing the collective value of privacy can provide philosophical clarity to the issues raised by data mining. Not only does a notion of collective privacy aptly address many of the ethical

concerns, it also accounts for the intuitive connections between privacy and data mining applications. Further analysis of the role that businesses using personal information play in shaping our collective view of privacy may lead to valuable policy changes that address the social duties that these businesses may have.

Chapter 1:

Questioning our Current State of Privacy

In 1999 Scott McNealy, former Chairman and CEO of Sun Microsystems Inc., responded to criticism over his company's newest networking software by famously stating, "You already have zero privacy – get over it."[1] When asked in an interview to comment on this statement he replied with the following,

> "The point I was making was someone already has your medical records. Someone has my dental records. Someone has my financial records. Someone knows just about everything about me. Gang, do you refute my statement? Visa knows what you bought. You have no privacy. Get over it."[2]

As our society becomes increasingly linked with advanced communications and information sharing across diverse sectors, one may well question whether or not we have lost what we traditionally conceived of as privacy. Personal data[3] is now collected, stored, and aggregated as never before. As innovative information technologies emerge, it is often difficult to determine whether our privacy has been affected.

Our networked economy and the fact that personal information has become a commodity have also renewed philosophical discussions on what "privacy" really means.

---

[1] Scott McNealy made this comment at a 1999 Sun Microsystems news conference, during the unveiling of the company's latest software; a new product with great networking abilities, called Jini. (Widely discussed on the internet and cited in many articles including: Austin, Lisa. 2003. Privacy and the question of technology. *Law and Philosophy,* **22:** 119-166.

[2] On the Record: Scott McNealy. 2003, September 14. *San Francisco Chronicle.* Retrieved January 14, 2006 from: www.sfgate.com/cgi-bin/article.cgi?file=/chronicle/archive/2003/09/14/BU141353.DTL&type=business

[3] The term "data" can take a singular or plural form. In the plural form, it refers to the plural of "datum". For example, "The collected data *are* further analyzed." "Data" can also be used in the singular form to refer to the collected data as a distinct collection. For example, "The collected data *is* further analyzed." In different situations one way of using the term often makes more sense than the other – i.e. whether one is referring to the pieces of data or to the collection. Thus both senses of the term are commonly used without undue confusion and both senses will be used in this thesis.

Throughout history, advances in technology have sparked some of the most renowned debates on the nature of privacy. In 1890, Samuel D. Warren and Louis D. Brandeis were reacting to the introduction of "instantaneous photography" when they famously argued that privacy is "the right to be let alone".[4] More recently, we have seen a shift from the Big Brother, Panopticon concerns of the 1970s when technology enabled the creation of large government databanks[5], to more pervasive surveillance concerns of the 1990s involving technologies such as Caller ID[6], workplace email monitoring[7], and closed circuit television cameras[8]. Today, concerns are being raised over subtle and ubiquitous information technologies such as data mining[9] that collect, store and use personal data as never before. We are truly becoming a monitored society, but is our "privacy" disappearing?

One way to explore connections between privacy and technology is to see whether accepted definitions of privacy adequately address the ethical issues associated with new technologies. In other words: Can the concerns raised by the technology be properly articulated as privacy concerns within the context of an accepted definition of privacy?

---

[4] Warren, Samuel D., and Louis D. Brandeis. 1890. The right to privacy [The implicit made explicit]: In, Schoeman, Ferdinand (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*. (pp. 75-103) New York: Cambridge University Press.

[5] Bennett, Colin J. 2001. Cookies, web bugs, webcams and cue cats: Patterns of surveillance on the world wide web. *Ethics and Information Technology* **3**: 197–210: p.198.

[6] *See for example*: Case, Donald O. 1998. The ethics of caller identification services. *Journal of Information Ethics*. **7**(1): 24-35.

[7] *See for example*: Sipior, J.C., B.T. Ward and S.M. Rainone. 1998. Ethical management of employee email privacy. *Information Systems Management*, **15**(1): 41–47.

[8] *See for example*: Goold, Benjamin J. 2002. Privacy rights and public spaces: CCTV and the problem of the unobservable observer. *Criminal Justice Ethics*, **21**(1): 21-27.

[9] *See for example*: Danna, Anthony, and Oscar H. Gandy Jr. 2002. All that glitters is not gold: digging beneath the surface of data mining. *Journal of Business Ethics*, 40(4): 373-386.

For example many privacy conceptions effectively explain why the information in medical records is properly considered "private" information, but it would be hard to use these conceptions to argue that a movie theater is a "private place" where "private actions" can take place.[10]

When privacy definitions do not provide a good context for articulating the ethical issues associated with a new technology then it is important to determine the reasons for this. It could mean that our privacy has already been weakened or lessened; we have already lost privacy. If this is the case our current conceptions of privacy might be lacking; perhaps the traditional ways of conceiving privacy are now inadequate in our technological society. It might be hard to identify this reduced state of privacy if the technology in question has caused a slow and subtle erosion. A careful analysis would need to be conducted to access these new threats to privacy and we would need to amend our conceptions of privacy or the ways that privacy functions in our laws and policies in order to fully address our technological concerns.

Alternately, it may be that technology is raising different issues that cannot and should not be grouped under a privacy heading. In this situation, addressing all of these matters as "privacy concerns" may cause confusion and lead to inadequate discussions not only of the issues at hand, but of privacy concerns in general.

Thus, it is through this method of applying an accepted definition of privacy to an ethically complex information technology that one can gain insight into our current state of privacy and discover whether the use of the technology really does pose a threat. In

---

[10] I realize that the distinction between the private and the public is contentious but this example is just meant to show how privacy definitions function normatively.

this thesis, I will follow this method to explore the role that privacy plays in the ethical analysis of data mining.

Data mining[11] is a technology that is now ubiquitous, and although it can appear benign, a closer examination reveals complicated underlying ethical concerns. I focus on data mining applications within the business sector, where large amounts of personal information are collected both online and offline. Data mining is a method that uses mathematical algorithms to extract implicit, previously unknown connections and patterns from large databases.[12] This technology enables businesses to sort consumers into various profiles which can lead to de-individualization.[13] Consumers who match the profiles deemed to be less profitable to a business may be denied incentives or services, or may be pressured to quietly end business with the organization. In the short term, this may lead to negative consequences for certain consumers. In the long term, this practice may cause serious societal harm if the profiles pervade our social consciousness and promote discrimination and de-valuing of certain groups of people.

Data mining is a complex topic as it relies on largely public, non-intimate and anonymous information that is used within the business sector – a sector with its own goals and ends. Sifting through the ethical concerns of data mining can be difficult.

---

[11] This short section on data mining is meant to serve as an introduction. In the following chapters, I will say much more about what data mining is, how it works, and what the ethical implications are.

[12] Fayyad, U., G. Piatesky-Shapiro and P. Symth. 1996. Discovery in databases: An overview. In G. Piatestsky-Shapiro and W.J. Frawley (eds.), *Knowledge Discovery in Databases*. Menlo Park, Cal, Cambridge, Mass/London: AAAI Press/MIT Press.

[13] "De-individualization" is a term coined by Anton Vedder (reference to follow). He defines de-individualization as, "a tendency of judging and treating persons on the basis of group characteristics instead of on their own characteristics and merits." The reasons for using this term will become apparent later in the thesis, during the ethical analysis of data mining. Vedder, Anton H. 1999. KDD: The challenge to individualism. *Ethics and Information Technology*, **1**: 275-281: p. 275.

Some critics[14] argue that the focus should shift away from privacy to other ethical frameworks such as discrimination, human rights violations and informed consent, as these might provide a more accurate and useful view of the problems. However, there seems to be an intuitive and pragmatic connection between privacy and technologies that collect, aggregate and use our personal information. I intend to show that acknowledging the relationship between the social value of privacy and information technologies can provide philosophical clarity to the complex ethical issues of data mining.[15] The policy implications that follow from this analysis are relevant not only to data mining applications but to other information technologies as well.

This thesis explores the connection between privacy and data mining but this connection is not an isolated concern. The larger commentary on information technologies and the nature and value of privacy in our society relates to many contemporary and traditional discussions of privacy. It touches on classic debates such as the distinction between the public and the private; capitalism and the social good; and the role of government regulation within the market economy. It also contributes to the exploration of our current state of privacy post-9/11 where we have seen some disturbing uses of technology and some disturbing justifications for privacy invasions.[16]

---

[14] *See for example*: Gandy, Oscar H. Jr. 2000. Exploring identity and identification in cyberspace. *Notre Dame Journal of Law, Ethics, and Public Policy*. **14**(2): 1085-1111.

[15] Please note that I am not setting out to redefine privacy or reject our traditional conceptions of privacy. By applying an accepted notion of privacy to the ethical concerns raised by data mining, I intend to draw attention to the way that privacy has a collective nature that is often overlooked. The importance of privacy as a social value has not been properly acknowledged in our laws and policies. I will develop the argument that information technologies are highlighting that this social value of privacy needs to be recognized if our privacy is to be properly secured. (Yet our traditional conceptions of privacy do not need to be abandoned in light of this).

[16] For example: earlier this year outrage was sparked by the discovery that the U.S. government was secretly compiling the records of everyday phone calls made by American citizens. Some saw this as an unacceptable attack on privacy while others felt it was a necessary part of the hunt for terrorists and that it

To start this analysis, I will clarify the definition of privacy that I will use throughout this thesis, highlighting its importance and relevance in some of the major philosophical debates on the nature of privacy.[17] In Chapter 2, I describe what data mining is, how it works, and what the advantages for businesses and consumers are. In Chapter 3, I discuss the immediate and long-term negative consequences that are associated with certain data mining applications. In Chapter 4, I examine whether my working definition of privacy provides a good context for articulating the ethical issues associated with this technology and what this analysis says about the normative function of privacy in our society. Relying on evidence that privacy has a wider social value than is commonly recognized, in Chapter 5 I intend to show that acknowledging the collective nature of privacy clarifies the ethical concerns raised by data mining and presents pragmatic and intuitive policy suggestions.

Privacy as Limited Access to the Person

I have stated that I will analyze the relationship between data mining and privacy by applying an accepted definition of privacy to the ethical issues associated with certain applications of this technology. The privacy definition that I will use is based on the claim that privacy involves limited access to the person; a view that is summarized by

---

did not constitute a worrisome invasion of privacy. *See for example: The Globe and Mail.* Sunday, May 12, 2006. Story: Huge database of phone calls a hidden trove of behaviors: Section A11.

[17] In this section it will become clear why I have chosen this definition of privacy to work with. I will show that other definitions of privacy could not contribute as much to the analysis of the relationship between technology and privacy.

Ferdinand Schoeman[18], developed and elaborated by Ruth Gavison[19], and amended from a feminist perspective by Anita Allen[20].

As I explain this conception of privacy, its relevance and influence within the philosophical literature will become clear. We will see how it addresses many of the conceptual difficulties encountered by other philosophers who have written on the nature and value of privacy; and how it conforms to our intuitions of privacy – acknowledging the necessary role of privacy for personal autonomy and situating privacy within a free and egalitarian society. My discussion here is not meant to be exhaustive; it is meant to give the reader an idea of some of the different ways that privacy has been conceptualized and why I have chosen privacy as "limited access to the person" as my working definition.[21]

Ferdinand Schoeman is the editor of one of the most comprehensive philosophical texts on privacy, *Philosophical Dimensions of Privacy: An Anthology.*[22] In the first chapter of this book, and in a corresponding paper[23], he summarizes some of the most

---

[18] Schoeman, Ferdinand. 1984. *Privacy: Philosophical dimensions*, American Philosophical Quarterly. **21**(3): 199-213. *And:* Schoeman, Ferdinand (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*. New York: Cambridge University Press.

[19] Gavison, Ruth. 1980. Privacy and the limits of the law. In, Schoeman, Ferdinand (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*. (pp. 346-402). New York: Cambridge University Press.

[20] Allen, Anita L. 1984. *Women and their privacy: What's at stake?* In Carol C. Gould (ed.), Beyond Domination. Totawa: Rowman & Allanheld.

[21] For more comprehensive reviews of the privacy literature, see for example: Schoeman, Ferdinand (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*; Allen, Anita. 2003. Privacy. In Hugh LaFollette (ed.). *The Oxford Handbook of Practical Ethics*. Oxford: Oxford University Press.; DeCew, Judith Wagner. 1997. *In Pursuit of Privacy: Law, Ethics, and the Rise of Technology*. Ithaca: Cornell University Press.

[22] Schoeman, Ferdinand (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*.

[23] Schoeman, Ferdinand. 1984. *Privacy: Philosophical dimensions.*

influential philosophical papers on the nature of privacy. He claims that proposed definitions of privacy can be separated into three categories. Briefly these categories are privacy as a right; privacy as a measure of control over information about oneself, personal intimacies, and sensory access; and finally privacy as limited access to the person.[24]

Schoeman rejects the notion of privacy as a right because he argues that it presupposes the moral status of privacy. It does not tell us why privacy is important to protect and it does not differentiate between the amount of privacy one has and the claim or right to privacy.[25] He also rejects the claim that privacy is a measure of control over information about oneself, personal intimacies, and sensory access. He believes this conception is vulnerable to counter examples. For instance: a person who freely discloses her personal information is exhibiting control over this disclosure but has no privacy. Thus it is important to distinguish whether someone has control over information and when someone is in a state of privacy.[26]

Schoeman accepts the third definitional category which is the view that privacy "is a state or condition of limited access to a person."[27] He claims that this characterization of privacy allows for discussion of whether privacy is a desirable state, it allows for some contemporary ethical issues relating to personal autonomy to be viewed,

---

[24] Schoeman, Ferdinand. 1984. *Privacy: Philosophical dimensions*: p. 199; *and* Schoeman, Ferdinand (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*: pp. 2-3.

[25] Schoeman, Ferdinand (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*: p. 3.

[26] Schoeman, Ferdinand. 1984. *Privacy: Philosophical dimensions:* p. 199.

[27] Schoeman, Ferdinand (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*: p. 3.

arguably, as privacy issues, and it allows for a distinction between loss of privacy discussions and right to privacy discussions.

Privacy as "limited access to a person" therefore appears to overcome the problems that the first two categories faced and it looks as though it may account for some of the intuitive connections between autonomy and privacy.[28] Schoeman explains that this definition also distinguishes between whether there are privacy "rights" such as moral rights or legal rights, and whether one has lost privacy. Based on Schoeman's expertise and his analysis of the philosophical literature on privacy, privacy as "limited access to the person" looks to be a very promising conception. However, from this brief description it is hard to know what "limited access" would mean in different contexts and how it would be achieved. More needs to be said about this definition before it can be fully appreciated.

One of the major proponents of this conception of privacy is Ruth Gavison.[29] She too rejects the notion of privacy as a form of control over information. She states that at one time we might criticize someone who chooses privacy and at another time we may criticize someone who does not choose privacy.[30] In other words, we may value or criticize the exercise of choice or autonomy apart from the value of privacy.

Gavison's thesis is that secrecy, anonymity, and solitude are all distinct yet inter-related components of privacy conceived as limited access to the person. Privacy can be lost due to one of these aspects, two aspects in combination, or all three in concert.

---

[28] *Ibid.*, p. 3.

[29] Gavison, Ruth. 1980. Privacy and the limits of the law.

[30] *Ibid.*, p. 350.

Firstly, secrecy involves limited access to information about a person.[31] Gavison states that a loss of privacy can occur when others gain information about an individual.[32] She notes that this component of privacy is not a derivate right reducible to other rights or claims, as other authors have argued.[33] For example, large government databanks, excessive information gathering during job interviews, and the publishing of love letters without the author's consent are all situations that involve a loss of privacy but do not involve any so-called derivative violations such as falsification, intrusion, trespass, et cetera.[34] Thus secrecy is properly an independent component of privacy, as it is privacy and not a derivative right, that is "related to the amount of information known about an individual."[35]

Anonymity is the second component and involves limited access to a person through limited attention paid to that person. Gavison writes,

> "An individual always loses privacy when he becomes the subject of attention. This will be true whether the attention is conscious and purposeful, or inadvertent."[36]

---

[31] The term "secrecy" here is not being used in the way that we might typically think. Sissela Bok (reference to follow) has argued that secrecy and privacy are distinct concepts and secrecy can be best understood as 'intentional concealment'. It is clear however that Gavison is using the term 'secrecy' to mean something like the contemporary notion of "informational privacy". I will continue to use Gavison's terminology for this thesis but please note that "secrecy" in this context refers more to the concept of "informational privacy". (*See*: Bok, Sissela. 1983. *Secrets: On the Ethics of Concealment and Revelation*. New York: Pantheon Books.).

[32] Gavison, Ruth. 1980. Privacy and the limits of the law: p. 351.

[33] Thomson, Judith Jarvis. 1975. The right to privacy. In, Schoeman, Ferdinand (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*. (pp. 272-289). New York: Cambridge University Press. (I will return to Thomson's argument and Gavison's rebuttal later in this chapter).

[34] Gavison, Ruth. 1980. Privacy and the limits of the law: p. 351.

[35] *Ibid.*, p. 351.

[36] *Ibid.*, p. 353.

Gavison states that losses of privacy occur when attention is paid to someone in a direct way such as following him, listening to him, observing him, et cetera. Other types of attention, such as discussing, imagining, and thinking about an individual, relate to privacy in an indirect way when these activities lead to greater interest and more attention paid to the individual.[37]

The final component of Gavison's conception of privacy is solitude, as solitude involves limited physical access to a person. Examples include such situations as Peeping Toms, and moving from your own office into a shared office. Gavison writes, "the essence of the complaint is not that more information about us has been acquired, nor that more attention has been drawn to us, but that our spatial aloneness has been diminished."[38]

These components are irreducible but they are also coherent. Gavison discusses some examples that explain how the elements may coexist, such as when a psychiatrist conducts therapy.[39] The psychiatrist gains information about her patient, pays attention to her, and sits near her; which involves all three components of privacy, yet none of the components implies the other two. Another example is of police surveillance: criminals might not be very persuasive if they claim their right to informational privacy was violated when police learned about their criminal plans.[40] However the criminals may have a better case if they claim that excess attention has been paid to them through constant police surveillance. These situations show that different privacy concerns can be

---

[37] *Ibid.*, p. 354.

[38] *Ibid.*, p. 354.

[39] *Ibid.*, p. 355.

[40] *Ibid.*, p. 355.

present in the same situation and how the different components of privacy as limited access (secrecy, anonymity and solitude) are distinct and interrelated.

Gavison also explains what privacy is not. She states that when more than the notion of limited access to the person through the elements of secrecy, anonymity, and solitude are included in a definition of privacy, more is lost than gained. Thus Gavison rejects, among other things,

> "exposure to unpleasant noises, smells, and sights; prohibitions of such conduct as abortions, use of contraceptives, and 'unnatural' sexual intercourse; insulting, harassing, or persecuting behavior; presenting individuals in a 'false light'; unsolicited mail and unwanted phone calls; regulation of the way familial obligations should be discharged; and commercial exploitation." [41]

Gavison claims that these are all examples of what have been dealt with as privacy invasions in the literature but they do not properly constitute distinct privacy concerns and thus lead to further confusion over what privacy really means. She states that the only way to catch all of these so-called losses of privacy is with an overarching definition such as the one given by Warren and Brandeis, who argued that privacy is the right to be let alone.[42] However definitions like this let in almost all conceivable instances of not being let alone into privacy discussions, even instances that have nothing to do with privacy invasions (such as requiring people to pay their taxes).[43]

---

[41] *Ibid.*, p. 356.

[42] Warren, Samuel D., and Louis D. Brandeis. 1890.

[43] Gavison, Ruth. 1980. Privacy and the limits of the law: p. 357.

She likewise rejects other broad definitions of privacy such as that of Edward Bloustein.[44] Bloustein claims that all violations of privacy are violations against human dignity. Gavison points out that there are many instances when human dignity is affronted and privacy is not involved (such as selling one's body to survive).[45]

Gavison also counters the argument that privacy properly understood as the right to be let alone involves non-interference of the state. Definitions of privacy that exclude the state also exclude claims that are not highly personal in nature, such as the wish to prevent your file from entering a data-bank.[46] Often these less personal privacy disputes rightly call for state involvement. Gavison writes, "A better way to deal with [non-interference] issues may be to treat them as involving questions of liberty, in which enforcement may raise difficult privacy issues."[47]

So far we have seen that Gavison's conception of privacy as limited access to the person accounts for many of the difficulties faced by other privacy definitions. Gavison's arguments are even more convincing when we consider the way her definition of privacy functions. She argues that, so conceived, privacy allows for autonomy; the promotion of mental health, human relations, and liberty of action; and the freedom from ridicule, censure, and physical access.[48] Limited access to a person based on the dimensions of secrecy, anonymity, and solitude allow for an individual to create an independent identity

---

[44] Bloustein, Edward J. 1964. Privacy as an aspect of human dignity: An answer to Dean Prosser. In, Schoeman, Ferdinand (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*. (pp. 156-202) New York: Cambridge University Press.

[45] Gavison, Ruth. 1980. Privacy and the limits of the law: p. 357.

[46] *Ibid.*, p. 358.

[47] *Ibid.*, p. 358.

[48] *Ibid.*

apart from possible hostile reactions of others. Such an identity is necessary in order for an individual to become a fully realized person. Gavison writes,

> "We desire a society in which individuals can grow, maintain their mental health and autonomy, create and maintain human relations, and lead meaningful lives. The analysis above suggests that some privacy may therefore both indicate the existence of and contribute to a more pluralistic, tolerant society."[49]

Gavison's arguments therefore explain not only how privacy promotes individual autonomy but also why privacy is an important part of a society that protects individual freedom. In this way Gavison's approach is both positive and intuitively satisfying and it allows one to respond to conceptions that do not view privacy as an important social value. For example, Gavison's approach can be used to counter the claim that privacy is a "derivative" right that can be explained by reference to other rights, as Judith Jarvis Thomson argues.[50] Thomson's argument lacks normative power when we consider Gavison's claims of the importance of privacy in-and-of-itself to human autonomy and society as a whole.

Gavison's arguments also overcome the less intuitive "market value" approach to personal information that one may interpret James Rachels and Charles Fried as arguing.[51,52] According to Rachels and Fried, intimacy depends not on the character of the relationship but on its exclusive nature; privacy is properly understood as the ability to share certain information with some and keep it hidden from others (as Jeffrey H.

---

[49] *Ibid.*, p. 369.

[50] Thomson, Judith Jarvis. 1975. The right to privacy.

[51] James Rachels. 1975. Why privacy is important. In Schoeman, Ferdinand (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*. (pp. 290-299) New York: Cambridge University Press.

[52] Fried, Charles. 1968. Privacy [A moral analysis]. In, Schoeman, Ferdinand (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*. (pp. 203-222) New York: Cambridge University Press.

Reiman[53] points out in his critique of Rachels and Fried). This view presented by Rachels and Fried is not very satisfactory as it denies our deep feelings for the character of intimate relationships. As Reiman points out, "the fallacy in the Rachels-Fried view of intimacy is that it overlooks the fact that what constitutes intimacy is not merely the sharing of otherwise withheld information, but the context of caring which makes the sharing of personal information significant."[54] Gavison's view of privacy as limited access explains why privacy allows for intimacy in a more intuitive way. In her analysis, privacy allows for intimacy and personal growth by allowing the individual to create an identity made up of many different roles. Thus different presentations can be shown to different people at different times to allow for different relationships to develop. Privacy also allows the individual to keep to herself her inner thoughts and emotions that may influence or damage certain relationships. Thus privacy allows for intimacy and relationships through the promotion of autonomy and the creation of a unique self.

Finally Gavison's conception of privacy as limited access to the person counters the claim that privacy is not an important societal value because it has not been extensively protected by the law.[55] She explains why it is that legal protection for privacy has seemed lacking in the past and why there are fewer privacy issues brought before the courts than are actually occurring. She states that in many cases the victim may not even be aware that a privacy violation has occurred. Also it is often the case that privacy violations involve situations where further attention is not especially desired, particularly

---

[53] Reiman, Jeffrey H. 1976. Privacy, intimacy and personhood. In Schoeman, Ferdinand (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*. (pp. 300-316). New York: Cambridge University Press.

[54] *Ibid.*, p. 305.

[55] Gavison, Ruth. 1980. Privacy and the limits of the law: p. 377.

not legal proceedings which may not show consideration for the sensitive nature of the violation. As well, legal battles are often costly and lengthy. Gavison therefore concludes that the legal protection afforded to privacy issues is not an adequate reflection of the importance of privacy to society.[56] She argues that more explicit privacy protection should be implemented both to reflect its importance in society and as an educational tool for the public.

This analysis has shown that Gavison's concept of privacy as limited access to the person is intuitive and satisfying. It addresses many of the common privacy concerns raised by philosophers, it situates privacy within a free and egalitarian society, and it offers positive policy suggestions. These features have made this definition of privacy very influential. As we saw, in his review of the philosophical literature, privacy as limited access to the person is a conception that Ferdinand Schoeman endorses.[57]

Another philosopher who has been influenced by Gavison's definition of privacy is Anita Allen. In her article, *Women and their Privacy: What's at Stake?*[58] Allen agrees that privacy as limited access to the person properly situates privacy as an important societal concern. Allen writes,

> "Privacy, understood as secrecy, solitude, and anonymity, promotes relaxation and intimacy by removing persons from view; it relieves pressures to conform; it promotes freedom of thought and (hence) of action; it promotes inventiveness and creativity. Privacy also limits unfavorable social responses to a person's actions, history, or plans, including public ridicule, moral censure, unfavorable decisions, and the formation of unfavorable opinion."[59]

---

[56] *Ibid.*, p. 371.

[57] Schoeman, Ferdinand. 1984. *Privacy: Philosophical dimensions.*; *And:* Schoeman, Ferdinand (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology.*

[58] Allen, Anita L. 1984. Women and their privacy: What's at stake?

[59] *Ibid.*, p. 238.

Although Allen finds Gavison's arguments very convincing for explaining the nature and function of privacy, Allen feels that Gavison's definition lacks certain requirements that are necessary to protect the privacy of women. Allen takes each dimension of privacy (secrecy, anonymity, and solitude) and discusses various situations where women encounter privacy violations that men may not encounter. She discusses how, within the dimension of secrecy, women may lose informational privacy at the hands of employers, the state, and university professors, where such professionals may require them to provide more detailed and inappropriate information that would not be demanded of male counterparts. Allen states that anonymity in public is harder for women to maintain than it is for men. Women are constantly harassed in public and may be considered "fair game" if they venture out.[60] Allen states that it is especially hard for women to find solitude, or privacy through limited physical access, in the home. Solitude for women often comes at too high a price if the traditional roles of women as caretakers, mothers and wives would need to be yielded in order to pursue self-directed interests. As psychologist Jean Baker Miller points out in Allen's article, women are seeking more than "autonomy as it is designed for men."[61]

Thus all these privacy invasions come at a cost to women. Women's careers, livelihood, and emotional states; their quality of life in their social communities;[62] and their self-definition and personal development can all be affected by these instances of

---

[60] *Ibid.*, p. 241.

[61] *Ibid.*, p. 243.

[62] *Ibid.*, p. 241.

privacy invasions that are unique to them.[63] Allen argues that what women need to secure privacy is *freedom of choice*. She states that without liberty with respect to sex, childbearing, and marriage, women will not be able to secure privacy.[64] She writes, "Free to choose, women enjoy privacy in the sense of *limited control by others of matters affecting their sexual and familial life.*"[65] Thus Allen claims that in addition to the general privacy interests that Gavison's conception of privacy includes, namely limited access to the person through secrecy, anonymity and solitude, another facet is needed. Women require Gavison's general privacy protection as well as the protection of, "liberty interests, or interests in limiting their accessibility to others through non-regulation and non-control by others of their sexual and familial lives."[66]

As stated earlier, Gavison explicitly excludes in her definition of privacy the very interests that Allen is arguing to include. Gavison claims that issues such as prohibitions on abortion and contraception, harassment, and regulation of familial obligations are not privacy issues at all but are rather liberty of action issues. Gavison criticized overly large conceptions of privacy such as the right to be let alone which included so many instances of supposed privacy violations that it confused what privacy properly means. Allen however claims that Gavison's account of what should be excluded as a privacy violation is somewhat arbitrary. Allen feels that freedom of choice is a necessary part of a privacy definition and its inclusion does not diminish the philosophical clarity. She writes that a

---

[63] *Ibid.*, p. 243.

[64] *Ibid.*, p. 244.

[65] *Ibid.*, p. 245.

[66] *Ibid.*, p. 245.

loss of liberty and a loss of privacy often go hand-in-hand without one excluding the other. Allen concludes,

> "A legal right to privacy must protect all of privacy's dimensions – secrecy, anonymity, solitude, *and* choice – if women's privacy interests are to be adequately protected. A society which protects these interests shields women from low self-esteem, emotional distress, and loss of opportunities and income. Liberal ideals such as liberty, equality, equal opportunity, and personal productivity are more surely advanced where women are treated with respect and they are free from discriminatory legal institutions, social policy, and customs."[67]

Thus Allen and Gavison both claim that their conceptions of privacy are necessarily part of an egalitarian society. Both state that their conceptions of privacy are important to allow for fully realized development of the person and also for the promotion of a just and equal society.

One may wonder whether Allen's addition to Gavison's position is necessary or whether, as Gavison might claim, more has been lost than gained. For the purposes of this thesis, I will not resolve this issue directly; I will see how Gavison's definition and Allen's amendment relate to my initial questions of privacy and technology. Specifically, I will examine whether or not these conceptions address the concerns raised by data mining technology. I will look to see if there are unique issues that would be better addressed by Gavison's conception or if there are ways in which Allen's addition alleviates certain problems. If both these ways of conceiving privacy are lacking with respect to the ethical concerns raised by data mining then what are the reasons? Perhaps data mining is eroding our privacy. Perhaps data mining raises concerns that may appear to be privacy issues but may not be privacy issues at all.

---

[67] *Ibid.*, p. 248.

I have chosen to use privacy conceived of as "limited access to the person" for my analysis because it is such a rich and intuitively satisfying concept. We have seen that it accounts for the importance our society places on privacy because it emphasizes the necessary role of privacy in the formation of autonomous individuals and a free and equal society. It allows a straightforward way to identify when privacy has been lost and it counters many other less satisfying conceptions of privacy. These qualities are why this definition will be useful for analyzing the relationship between data mining and privacy and for examining the broader connection between privacy and information technologies. Before I begin this analysis, in the next chapter I will explain what data mining is, how it works, and some of its advantages for businesses and consumers.

Chapter 2:

Data Mining

In this chapter, I will give a brief summary of what data mining is and how it works. I will then explain how data mining differs significantly from traditional data analysis. I will also highlight the data mining advantages for businesses and for customers.

Data Mining and How it Works

Succinctly put, "Data mining is a technique used to discover hidden information in very large databases."[68] What most people refer to as "data mining" is technically "knowledge discovery in databases" (or KDD). Data mining is actually one step of knowledge discovery in data bases, though most people now use the term "data mining" to refer to the whole of the KDD process. The full definition of KDD is, "the non-trivial extraction of implicit, previously unknown, and potentially useful information from data."[69] Bart Custers identifies five different steps in the KDD process[70], though several authors[71] now group some of these steps together and discuss only three. For clarity I

---

[68] Custers, Bart. 2001. Data mining and group profiling on the internet. In Anton Vedder (ed). *Ethics and the Internet*. Antwerpen-Groningen-Oxford: Intersentia: p. 89.

[69] Frawley, W.J., G. Piatetsky-Shapiro and C.J. Matheus. 1991. Knowledge Discovery in Databases: An Overview. In G. Piatetsky-Shapiro and W.J. Frawley (eds), *Knowledge Discovery in Databases*. Menlo Park, Cal., Cambridge, Mass/London: AAAI Press/MIT Press.

[70] Custers, Bart. 2001. Data mining and group profiling on the internet.

[71] Vedder, Anton H. 2001. KDD, Privacy, Individuality and Fairness. In Richard A. Spinello and HermaT. Tavani (eds). *Readings in CyberEthics*, Jones and Bartlett.

will summarize the five different KDD steps as explained by Custers. Since "data mining" is now the common term for the whole KDD process, for the remainder of this thesis I will use the term "data mining" instead of KDD.

Before I proceed, I would like to point out that there are many different applications of data mining technology. Notably, there are numerous valuable data mining applications that mine information in large databases that do not contain any personal data. For example, "Nora" is a system that allows for data discovery in literary studies.[72] An extremely large database of literary work can now be searched for various overarching, complicated themes such as "sentimentalism". Previously, searching such a large number of works manually was essentially impossible to do. Data mining applications such as Nora do not involve personal data and thus do not raise ethical concerns. Applications such as these will not be the focus of this thesis. There are other data mining applications that do not involve personal data but do involve ethical dilemmas, such as those involving animals. Data mining applications can also be used for certain law enforcement and political purposes and this can lead to contentious issues. However, for the purposes of this thesis, I will focus on data mining applications involving personal data, mainly within the business sector as the issues raised there are some of the most controversial data mining issues today.

Some quick examples of the data mining applications that I will explain in more detail later are: grocery store client cards, websites that require users to login, and store coupons or raffles that ask for personal information. In each of these cases, personal

---

[72] Rueker, Stan. 2006, January 12. *Crystalizing the Process: the Clear Browser for Data Mining Humanities Computing*, Talk at the University of Victoria, Victoria, B.C.
(For more on Nora visit: www.noraproject.org.)

information is collected from customers or users. This information is stored in databases and linked to other personal information such as future store purchases or which websites are viewed. All of this linked information is aggregated with similar data from other customers and analyses are performed to see what the market trends are and what business policies would be most profitable. Again, I will explain these applications in more detail and give more examples later in this chapter.

Custers identifies five steps in the data mining process: data collection, data preparation, data mining, interpretation, and "determine actions".[73] Data collection can involve many different sources. Information can be collected off-line at the check-out through the use of "client cards" for example. Information can also be collected from customer questionnaires filled out manually at the store or online. Many websites require a "login" before the site can be accessed. This creates a situation where users can only view the site if they have filled out a questionnaire and provided data. These are more explicit ways to collect data, though the customer or user might not be aware of it. Data can also be collected more subtly through Web mining. Kosala, Blockeel and Neven[74] neatly summarize the three common ways to mine information from the Web:

> "1) Web content mining: investigating the content of documents in order to find relevant information.
>
> 2) Web structure mining: using the structure of the Web (i.e. the way in which different web pages are linked together) to find relevant information.

---

[73] Custers, Bart. 2001. Data mining and group profiling on the internet: p. 89.

[74] Kosala, R., H. Blockeel, and F. Neven. 2002. An Overview of web mining. In J. Meij (ed.). *Dealing with the Data Flood: Mining Data, Text and Multimedia*. Rotterdam: STT.

> 3) Web usage mining: using previously stored knowledge about the behavior of human users of the Web (for instance, how they navigated through the Web) to find relevant information."[75]

Thus content mining involves mining data appearing on web pages, whereas structure mining involves mining the data "regarding the hyperlink structure within and across web documents."[76] Some authors group these categories together as they state that the hyperlinks' data is more valuable when it is linked to contents.[77] Usage mining involves the monitoring of surfing behavior or the so-called clickstream data, tracking users online.[78]

There are many different methods used to collect surfing information online. One of the most commonly discussed techniques uses "cookies". van Wel and Royakkers write,

> "Cookies are small files that are placed on the hard disk of the web user during his browser session. The cookie will make sure that the web user's computer will be recognized and identified the next time the same web site is visited. Therefore, cookies can be used to track a user online, and enable the creation of a profile without him/her realizing it."[79]

Even though cookies are still discussed in the literature, Colin Bennett notes that they may be on their way out.[80] Along with various cookie blocking software now available, there may be new monitoring devices emerging. Bennett discusses the "Web Bug", a new tracking technology that consists of a practically invisible graphic on a web page or

---

[75] *Ibid.*, p. 484.

[76] van Wel, Lita., and L. Royakkers. Ethical Issues in Web Data Mining: p. 130.

[77] *Ibid.*

[78] Custers, Bart. 2001. Data mining and group profiling on the internet.

[79] van Wel, Lita, and Lambèr Royyakkers. 2004. Ethical issues in web data mining: p. 132.

[80] Bennett, Colin J. 2001. Cookies, web bugs, webcams and cue cats: Patterns of surveillance on the world wide web.

email message. The Web Bug tracks various information including, "the IP address of the computer that the Web Bug is sent to, the URL of the page the Web Bug comes from and the time it was viewed."[81] There are doubtless other devices available, and this sketch is not meant to be exhaustive. Rather it is meant to provide some examples of the various data collection methods used in data mining.

It is easy to see that there are differences in these data collection methods. The social ramifications of each method may be examined separately and many authors do this. However, it is also easily seen how many of these applications can work together. Consider the following example: Let's say, at a store, I am asked to give my email address to receive a special email coupon (this recently occurred when I visited a GAP clothing store – the rest of the example is a fabrication). I leave my name and my email and perhaps my address. Soon I receive my special email with a code to access my coupon. I click the link to get to the store's web page and enter my coupon code as well as some more information that I am asked to provide, perhaps my approximate age and number of people in my house and their age ranges. I am granted access and browse the site. I click on various product links and eventually use my coupon toward the purchase of an item. Of course I am assured that my privacy is important to this business and will not be sold or shared with third parties. This company now has my name, address (geographic location), email, the number and age ranges of people in my house as well as myself, the date and time I entered and left the site, the types of things I looked at that this company sells, and what I bought – it's price range, etc. – from this company. I can also be identified and more information added to my profile the next time I use my special code. Other information about my surfing behaviour might be collected if the site uses

[81] *Ibid.*, p. 209.

cookies or Web-Bugs. This example shows that many data collection methods can be used in tandem to collect more information than just one method could. It is for this reason that I will not be focusing on one specific data mining information collection method as varying techniques are used within the business sector.

After the collection of the data, there are still four steps left in the data mining process. The second step is the data preparation. In this stage the data is rearranged and ordered and sometimes aggregated, such as by postal code.[82]

Step three is the actual "Data Mining" step, from which the KDD process gets its everyday name. Simply put, data mining uses mathematical algorithms to discover connections and patterns hidden within large databases. In this section I will give a brief overview of what information is generated in the data mining step along with a few examples. Again, this explanation is not meant to be exhaustive. For an overview of the technical means of producing this information, such as neural networks, decision trees, market basket analysis, et cetera, see Danna and Gandy.[83]

In the data mining step, five types of information can be generated that are useful for marketing purposes. These include: associations, sequences, classifications, clusters and predictions.[84] I will give a brief explanation of each of these along with an example. Please note that these examples are meant to serve an explanatory role and therefore are not overly complicated. Information generated from data mining analysis can be quite complex and have far reaching consequences, which I will discuss later in this thesis.

---

[82] Custers, Bart. 2001. Data mining and group profiling on the internet.

[83] Danna, Anthony, and Oscar H.Gandy Jr. 2002. All that glitters is not gold: digging beneath the surface of data mining.

[84]Custers, Bart. 2001. Data mining and group profiling on the internet: p. 90.

"Associations are made when comparable events are connected to each other."[85] They entail events that are connected with no elapsed time. As an example, suppose it is discovered that seventy percent of internet airline ticket passengers are likely to purchase hotel accommodations at the time they purchase their flight. "Sequences [on the other hand] are connected successive events."[86] For example, perhaps it is the case that people are likely to buy the next book published by the same author of the book they have just purchased. Thus sequences involve elapsed time whereas associations do not.

The next two types of generated information, classification and clustering, are closely related. Custers notes that classification is most frequently used in data mining as it can identify customer preferences. He writes, "Classification is the examination of known groups to determine which characteristics can be used to identify or predict group membership: for example, which hyperlink on a particular Web site is clicked on most frequently and by what kinds of people."[87] Let us suppose that information is uncovered showing that parents typically click on family travel links. Here the known group is "parents" and the hyperlink is "family travel". After further analysis is done in the next steps of the KDD process, the company may use this information to assume that users identified as "parents" will be interested in "family travel" and prompt them or entice them accordingly. Clustering, on the other hand, involves discovering which groups exist within the data. Thus known characteristics are used to determine certain groups. Perhaps it is discovered that students are willing to travel farther than families with children. Here the known characteristic is "willing to travel farther" and the group

---

[85] *Ibid.*, p. 90.

[86] *Ibid.*, p. 91.

[87] *Ibid.*, p. 91.

discovered to exhibit this characteristic is "student" as opposed to "families with children."[88] After further analysis, the company may assume that those interested in extensive travel abroad are students or have student qualities – young, no children, etc. – and conduct business with them according to this assumption. In summary, classification involves the use of certain groups to determine the characteristics of the group, whereas clustering is the use of characteristics to determine which groups exhibit these characteristics.

"Predictions are time extrapolations of parameters," and Custers notes that predictions may provide the most valuable information to internet companies because this information will allow them to stay ahead of the game.[89] Predictions can be made on many parameters such as supply and demand, costs, wages, et cetera and can be simple or complex, for example involving the influences of the seasons.

After the data has been mined, the interpretation of the results occurs in step four. The results usually need to be converted from statistics into a more usable form. A selection of the results is also made at this stage, as to which information is most useful .[90]

The final step is the determination of what course of action the company will take given the information they now have. This step, "consists of determining corresponding actions – actions such as direct mailings to particular target groups, special offers to

---

[88] *Ibid.*, p. 91.

[89] *Ibid.*, p. 91.

[90] *Ibid.*, p. 92.

potential buyers, excluding people who almost certainly will not buy anything, et cetera."[91]

It is easy to see how the acquired information from each example in the data mining section may be interpreted and used for marketing purposes. Continuing with the examples used before: In the case of associations, businesses may offer joint accommodations with flight purchases to increase the sales of both. With information from sequence data, a book seller may offer to alert you when a new book by the author you just bought arrives. A company may use classification data to enable a website to group you into the class of "parent" if you click on a family travel site and this may lead to other promotions or enticements designed for parents. Based on clustering data that links overseas travel with students, one who identifies herself as interested in overseas travel may be notified of student travel incentives. With the use of predictive data, a company may start promoting various items at certain times of the year, whereas they had not done this before. These examples are meant to provide some ideas of how the data mining process works and what the corresponding actions of businesses may be based on. I will now explain some of the ways in which data mining is markedly different from traditional information retrieval methods.

The first and most important way data mining differs from traditional information retrieval methods is that whereas traditional methods involve posing an explicit question of the data and receiving an answer, data mining discovers patterns and trends within the data that are implicit. In traditional information retrieval, explicit questions are asked of the data. For example, a travel company could review the data they have collected to

---

[91] *Ibid.*, p. 92.

determine set hypotheses. They might try to find the average length of the vacation flights they sell or whether they sold more business flights to men or women this year compared to last year. They could then use this information, say, to promote flights of roughly the average length to those who inquire about vacations, and to develop an ad campaign to promote more business flights to women since this was the faster growing market. Thus traditional data retrieval involves asking set questions and receiving answers.

Data mining, on the other hand, involves the uncovering of implicit patterns and connections. Questions are not posed to the dataset, rather "answers" are uncovered. Tavani states, "Using data mining techniques it is possible to unearth patterns and relationships, which were previously unknown, and to use this 'new' information, i.e., new facts and relationships in the data to make decisions and forecasts."[92] For example, based on the data collected by the travel company, they might discover that people who booked more than eight business trips a year hardly ever booked vacation trips, and eighty percent of the people who book two consecutive annual vacations will book a third the following year. Thus the travel company may stop promoting vacations to frequent business customers and may offer incentives to vacationers to use their company more than once. The company set out to find relationships and patterns but it did not set out to find *these* specific relationships; rather the information emerged from the data. Thus, information retrieved using traditional methods is generally predictive in nature – the answers retrieved from the dataset will be based on the questions that are asked. On the other hand, there is a non-predictive aspect to information retrieved through data mining

[92] Tavani, H.T. 1999. Informational privacy, data mining, and the internet. *Ethics and Information Technology*, **1**: 137-145: p. 140.

analysis – the patterns and connections that are "revealed" are not based on established criteria because specific questions are not asked to the dataset. In this way, data mining allows for businesses to make use of "novel" information in the form of patterns and relationships that had not been previously recognized.

The explicit versus implicit difference is amplified by another difference, namely that data mining allows for processing extremely large data bases which were basically impossible to search through manually. Companies can now collect, store, and process data from one large database, called a "data warehouse". Data warehousing is not a requirement for data mining but it does enhance the potential returns. Traditionally data is stored in multiple databases. If information is to be shared, it needs to be retrieved across these databases. The limitations of traditional data storage mean that only specific data is retrieved and compared across the databases, again in an explicit manner. The large volume of data in the data warehouse allows for more extensive implicit data mining analysis. Companies can now collect and store all kinds of information that may have no apparent value. If not all of the data reveals useful information, it can be stored and analyzed at some other time. A well-known example of data warehousing in action is the WalMart retail chain. WalMart is often referred to as a "pioneer in data mining and data management."[93]

> "WalMart captures point-of-sale transactions from over 2 900 stores in six countries and continuously transmits this data to its massive 7.5 terabyte data warehouse. WalMart allows more than 3 500 suppliers to access data on their products and perform data analyses. These suppliers use this data to identify customer buying patterns at the store display level. They use

[93] Cavoukian, A. 1998. Data Mining: Staking a Claim on Your Privacy. *Information and Privacy Commissioner's Report*, Ontario, Canada: p. 8.

> this information to manage local store inventory and identify new merchandising opportunities."[94]

It is easy to see how data mining has changed the nature of information retrieval and data analysis. The analyses of WalMart and countless other businesses would not be possible using traditional methods.

Some of the data mining advantages to businesses may now seem obvious but they are worth stating explicitly. One of the major advantages to business is that data mining allows for construction of consumer or group profiles. Custers writes, "A group profile is a property or a collection of properties of a group of people."[95] As discussed earlier, identifying these groups involves classification (also called segmentation), determining certain attributes of a group, or clustering, determining which groups have certain attributes. These profiles are usually based on statistical analysis and thus are averages or probabilities. For example, a profile might be based on the finding that roughly eighty percent of university students buy a book online once every three months. The profiles used most often are therefore "nondistributive" – profiles whose properties are not necessarily true for every member of the group. (When a property or properties holds for every member of a group then this is called a distributive profile).

As we saw, there is a non-predictive aspect to information retrieved through data mining – instead of asking specific questions to the dataset and receiving specific answers, as was the traditional method used to retrieve information, patterns and connections "emerge" during data mining analysis. Thus, the analysis may reveal new or novel patterns or connections. Interestingly, the consumer profiles that are created using these

---

[94] *Ibid.*, p. 8.

[95] Custers, Bart. 2001. Data mining and group profiling on the internet: p. 92.

patterns and connections are very useful to businesses because profiles have a *predictive* nature. Before data mining takes place, the characteristics of a profile may be unknown. The analysis reveals relationships which become the make-up of the profile, and it is these relationships that have a non-predictive aspect; companies cannot predict which information will be important and which information will make up the profiles. Yet, once a profile is determined, a consumer who fits certain characteristics can be assigned to a certain profile with the assumption that the other features of the profile will apply. In other words, once profiles are created they can be used to predict how current and future customers will act within the marketplace.[96] In this way, group profiling allows companies to target specific groups rather than targeting individuals, which can be more costly. Group profiling also allows businesses to predict future customers and to tailor business dealings to the profile of the customer they are dealing with. For example, if I am identified as a "parent" by an online clothing company, I may be offered periodic updates on the children's line, whereas if I am identified as a "student" I may be offered periodic updates on the sale items. Custers also notes that, for businesses,

> "Group profiling is, in most cases, more useful than no profiling at all. Advertising with inadequately predefined target groups, like on television, is definitely less effective and less efficient than advertising only to interested and potentially interested customers."[97]

---

[96] Of course, predictions of how consumers will act are still "guesses"; data mining allows for advanced, calculated guesses, which are more profitable to businesses than the traditional ways of predicting customer behaviour.

[97] Vedder, A.H. KDD, Privacy, Individuality and Fairness: p. 94.

Profiling is also used to predict "default values". For example, a car company may find a certain profile that shows a preference for red cars. If someone viewing their Website matches this profile, he might be shown all red cars, with the option to change colours.[98]

Another data mining advantage for businesses is the use of the technology to detect fraud. Customer credit card fraud is a major worry for businesses, and data mining can be used to detect suspicious credit card activity. Data mining technology is also used to detect fraudulent employee practices.

The last advantage I will present is the use of data mining to "identify internal inefficiencies and then revamp operations."[99] A website, for example, can be monitored to see where customers are losing interest or it might be revealed that an application is taking too long and costing business.

It is clear that there are definite business advantages to using data mining technology; there are also advantages for the customer. One of the major advantages expounded by businesses is that data mining can lead to less unsolicited marketing. Instead of sending out mass advertising "junk" mail or "spam" emails, profiling allows for the targeting of specific groups who may want a certain products. Those who may genuinely be interested in a product are notified, while those who are most likely not interested in a product will not be. When van Wel and Royakkers interviewed professionals with businesses engaged in web-data mining they noted that, "some

---

[98] Custers, Bart. 2001. Data mining and group profiling on the internet.

[99] *Ibid.*, p. 111.

interviewees even foresee a possible occasional alliance between privacy protectors and data miners because of their mutual goal: less solicited marketing contacts."[100]

Another touted consumer advantage is the claim that data mining leads to, at least the appearance of, individualization. When interactions are tailored to specific profiles, consumers who match their assumed profile will feel that the interaction is more personalized. As van Wel and Royakkers point out, "Most customers like to be recognized and treated as a special customer."[101] For example, when you are searching certain online bookstores, other books may be presented in the margins. These additional books relate to your search information or other information you have provided. You may be very thankful for these suggestions when they match well to your interests. You might feel that this feature saved you time and helped you out. Thus data mining advocates state that it is beneficial to the customer when usage interaction is analyzed so that personalization is optimized.

In addition, customer incentives are frequently offered to those who volunteer certain personal data. Coupons or special offers may accompany leaving your name and address or other information. There are often chances to win prizes at grocery stores and online if certain information is supplied. Discounts are usually available if you use a loyalty card at certain businesses. Particular web sites grant exclusive access only to members (those who have filled in the appropriate information) and sometimes will ship for free or offer other bonuses if additional information is provided. Thus it can be argued that some of the returns of data mining are passed on to the customer or user. It is

---

[100] van Wel, Lita, and Lambèr Royyakkers. 2004. Ethical issues in web data mining: p. 135.

[101] *Ibid.*, p. 136.

so advantageous to businesses to collect data and process it that they are willing to make it worth your while to provide it.

In conclusion, we have seen what data mining is and how it works as well as some of the business and consumer advantages. In the next chapter, I will explain some of the social implications of data mining. In Chapter 4, I will connect data mining back to the privacy discussion in Chapter 1, and analyze how privacy and data mining are related.

Chapter 3:

The Ethical Concerns Associated with Data Mining

In the last chapter, I explained that data mining is a technique used to uncover implicit information hidden within large databases. The major advantages from both business and consumer perspectives are due to clustering or classification techniques, which sort consumers based on certain properties or a collection of properties. This enables businesses to create group profiles, which are very useful because of their predictive nature. Once a profile is determined, a consumer who fits certain characteristics can be assigned to a certain profile with the assumption that the other features of the profile will apply. Group profiling therefore allows companies to target specific groups rather than targeting individuals, which can be more costly. Group profiling also allows businesses to predict future customers and to tailor their business dealings to the profile of the customer they are dealing with. Data mining advantages for consumers may include less unsolicited marketing, incentives such as coupons in exchange for their personal information, and the appearance of a more personalized interaction with the business.

As advantageous as data mining may appear, there are complicated underlying ethical concerns associated with its use. Part of the reason that an ethical analysis of data mining is so complex is because, at first glance, it looks as though data mining does not raise any privacy concerns. In this chapter, I will explain why many data mining advocates claim that data mining does not threaten privacy. Interestingly, many data mining critics agree that privacy is not a central issue with respect to data mining but they

do not agree with the conclusion that the technology is ethically sound. In the second part of this chapter I will explain some of the serious social consequences associated with data mining, and I will discuss the reasons that critics disagree on the best way to articulate these problems and work for change. In Chapter 4, I will examine the connections between these negative consequences associated with data mining and the concept of privacy as limited access to the person – the privacy framework I outlined in Chapter 1.

The Apparent Difficulties with Connecting Data Mining to Privacy Concerns

At first glance, it may look as though data mining does not give rise to controversial privacy issues and the connection between privacy and data mining is a contentious one. Major defenses of data mining include the following arguments: information is collected in public and it is not intimate in nature; most information is freely given by consumers or users; and for the most part, the information is used anonymously.[102] These arguments are based on the assumption that data mining does not raise any (unique) privacy concerns and therefore it is an ethically benign technology. Understanding the arguments that claim that data mining does not threaten privacy will highlight the controversial nature of this debate and will better situate the following ethical analysis of data mining applications.

Firstly, the type and collection of information used in data mining analysis does not seem to evoke typical privacy concerns. At the grocery store, everyone can see what I have in my cart. As I place my groceries on the conveyor belt I do not try to shield them

[102] van Wel, Lita, and Lambèr Royyakkers. 2004. Ethical issues in web data mining.

from the eyes of others.[103] Since public shopping is usually not a secretive process, is it of any significance if the store records what I have purchased? What if the clerk at an electronics shop codes in some information about me along with my purchase, such as my gender and approximate age; again, this is public information that anyone at the store has access to. If I fill out a contest questionnaire at a store or if I "register" to use a web site online, I am choosing to provide information about myself.[104] Furthermore, if all this information is then de-identified such that all that is known is that a woman in this age range, living in roughly this neighbourhood, bought this and that, how has my privacy been affected? For the most part, no one could look into the data warehouse and know this information was really about me.

When we think of traditional controversial privacy issues we often think of intimate identifying information collected in private that one would wish to keep private. We might think of people who have been publicly and detrimentally exposed as being homosexual. Surely their privacy was violated when this intimate, personal information was collected and revealed against their will. We might think about the privacy of medical records as they identify intimate information about us that we would not want others to know. We also might think about the privacy of our bodies and the right we feel to determine who knows intimate details about us and who does not. Because control

---

[103] This is true for the most part. Sensitive grocery store items such as tampons or condoms may commonly be purchased in a more furtive way than say tomatoes or oatmeal. My point is that we do not usually conduct our "public" business in a covert fashion. This may imply that we do not care who knows what about our public business but it could also mean that there is a normative assumption that even our public interactions will be respected as somewhat private. I will analyze this point toward the end of my thesis.

[104] This is a common argument by many data mining advocates but many critics claim that consumers are often not properly informed before they consent to providing personal information. Privacy policies are often ambiguous; information can be collected in extremely subtle ways; and special offers or incentives to provide information may raise questions of coercion. I will return to these issues involving informed consent of the customer toward the end of this chapter.

over intimate, identifying information collected in private seems to be at the heart of privacy claims, many data mining advocates state that the technology does not raise any privacy concerns.[105] After all, data mining usually involves information that is freely given or collected in public, is non-intimate in nature, and is often used anonymously.

Data mining advocates also state that if privacy is a concern, pre-existing moral codes and our current laws sufficiently protect the personal privacy of the consumer.[106] Furthermore many online and offline businesses maintain that their privacy policies adequately acknowledge the privacy concerns of the customer or user.[107] Often it is expressly stated that customer privacy is very important and the company will not rent or sell your information to third parties. For example, the privacy statement on the WalMart website states,

> "The security of your personal information is very important to us. We never sell or rent your personal information to any third parties under any circumstances. We value your trust very highly, and will work to protect the security and privacy of any personal information you provide to us and will only use it as we have described in our Privacy Policy."[108,109]

The WalMart privacy policy, as with the privacy policies of most online businesses, explains in detail when and how personal information is collected. It states straightforwardly that along with collecting the freely given identifying information (such

---

[105] van Wel, Lita, and Lambèr Royyakkers. 2004. Ethical issues in web data mining: p. 134.

[106] *Ibid.*, p. 134.

[107] *Ibid.*, p. 134.

[108] The comments I made in the footnote 104 apply here too. This privacy policy of WalMart highlights the loaded terms often contained within these privacy policies. For example, what does "personal information" mean in this context? What "personal information" will they "work to protect"? Does it mean all the information you have provided to the company or just the information that has been deemed "personal" by the company? I will discuss this in more detail later in the chapter.

[109] *WalMart Privacy Policy*. Retrieved March 4, 2006 from http://www.walmart.com/catalog/catalog.gsp?cat=538446

as your name and address, etc.), "Web beacons" (something like, if not the same as, Web bugs[110]) and cookies are used to monitor online behaviour. It notes, "The purpose of the cookie is to identify you when you visit this Site so we can enhance and customize your online shopping experience."[111] The policy states that the user can block cookies, however purchases cannot be made nor are "special features" available to those who block the cookies. As well, the site explains that online information is monitored and tracked by third parties but the information in question is anonymous and used for marketing purposes – to measure the performance of and the response to marketing efforts. Among other reasons, such as fraud prevention, WalMart claims that the collection and use of customer information enhances your shopping experience as it allows the store to, "Communicate great values and featured items to you. Find and stock the products you want. Customize your shopping experience."[112] Thus the WalMart privacy policy is quite frank about what information is collected and what it is used for. It appears that most of this information is public information and seems largely harmless and anonymous. Furthermore, the claim is that it is collected with you, the consumer in mind, to enhance and customize your shopping experience. All consumers and users of the WalMart website can easily click the link to this privacy policy. If a consumer does not agree with the collection and use of information, then she is free to shop elsewhere. Again, everyone can read this privacy policy and decide whether or not they wish to continue to do business with the company, or not.

---

[110] Bennett, Colin J. 2001. Cookies, web bugs, webcams and cue cats: Patterns of surveillance on the world wide web.

[111] *WalMart Privacy Policy*. Retrieved March 4, 2006 from http://www.walmart.com/catalog/catalog.gsp?cat=538446

[112] *Ibid.*

From this discussion it looks as though data mining involves 1) the collection of public, non-intimate, often de-identified data, 2) using collection methods that are explicitly stated, 3) for the purposes of marketing and enhancing the shopping experience of the customer – what are the controversial issues? Many data mining advocates claim that there is nothing of ethical concern about business policies that use data mining analysis because data mining does not affect our privacy.[113] The argument can be roughly sketched in this way: Privacy properly involves intimate, identifying information that is collected in private. For the most part, it is information that we would like to keep private.[114] Data mining involves the collection and use of personal information that is mostly public, non-intimate, freely given and de-identified. The collection and use of this kind of personal information is therefore not a privacy issue and is largely ethically benign. The conclusion is that, since the collection and use of personal information does not cause privacy concerns, the technology is ethically benign as well.

It is easy to see from the way that I have presented the argument there are obvious places for criticism. For instance, it does not necessarily follow that if there are no privacy concerns stemming from the collection and use of personal information then this information is ethically benign. It also does not follow that if the collection and use of the information is ethically benign then the technology is too. However, I am not trying to set up a straw man argument. The ethical analysis of data mining is more difficult than it may appear. Any critique must address the apparent non-controversial nature of personal

---

[113] van Wel, Lita, and Lambèr Royyakkers. 2004. Ethical issues in web data mining: p. 134.

[114] As I explained in chapter 1, privacy allows for autonomy; the promotion of mental health, human relations, and liberty of action; and the freedom from ridicule, censure, and physical access. Limited access to a person based on the dimensions of secrecy, anonymity, and solitude allow for an individual to create an independent identity apart from possible hostile reactions of others. Such an identity is necessary in order for an individual to become a fully realized person. Thus we are properly concerned about losses of privacy that occur when others gain access to sensitive information about an individual.

information that is largely public, appears to be freely given, and is usually de-identified – this is not an easy task. While many critics agree that there are troubling trends associated with certain data mining applications, they disagree on which ethical framework can best address these problems. Some critics[115] maintain that our privacy is affected by data mining policies and argue that privacy does have a place in the ethical analysis of data mining, while other critics[116] accept that privacy is not so centrally involved. Instead, they argue that the focus should shift away from privacy to other ethical frameworks such as discrimination, human rights violations, and a lack of informed consent, as these might provide a more complete view of the problems and provide a better position from which to work for social change. The following analysis will show why ethical assessments of data mining have been quite divergent.

Ethical Analysis of Data Mining Applications

The major problem that many critics of data mining focus on relates to the same feature of data mining that many advocates consider to be its major asset, namely group profiling. Remember that data mining can use clustering or classification techniques to group consumers based on certain properties or a collection of properties. These profiles are usually based on statistical analysis and thus are averages or probabilities.[117] The profiles used most often are therefore "nondistributive" profiles or profiles whose properties do not necessarily apply to every member of the group.

---

[115] *See for example*: Tavani, Herman T. 1999. KDD, data mining, and the challenge of normative privacy. *Ethics and information technology*. **1:** 265-273; and Austin, Lisa. 2002. *Privacy and the question of technology*. Law and Philosophy, **22:** 119-166.

[116] *See for example*: Gandy, Oscar H. Jr. 2000. Exploring identity and identification in cyberspace. *Notre Dame Journal of Law, Ethics, and Public Policy*. **14**(2): 1085-1111.

[117] Custers, Bart. 2001. Data mining and group profiling on the internet.

Group profiles are very useful to businesses because of their predictive nature. Once a profile is determined, a consumer who fits certain characteristics can be assigned to a certain profile with the assumption that the other features of the profile will apply. Thus group profiling can allow companies to target specific groups rather than targeting individuals, which can be more costly. Group profiling also allows businesses to predict who will be future customers and to tailor their business dealings to the profile of the customer they are dealing with. Group profiling can also make customers feel that the interaction was customized to them specifically and thus was more personalized and rewarding.

The ethical problems associated with group profiling can be roughly divided into: questionable business practices, epistemological concerns, and larger social costs. These divisions are used here as a tool to help to clarify the underlying issues. Each category represents a separate concern but not an isolated concern. Problems in one category can compound the ethical considerations in another category. Both the questionable business practices and the epistemological concerns contribute to the larger social costs of data mining. Because of this, the full relevance of the questionable business practices and the epistemological concerns will only become apparent in the section where the larger social costs are explained and more extensive examples are given.

Data mining technology has been linked to an increase in questionable business practices.[118] The practices that I will discuss are price discrimination, customer firing,

---

[118] Danna, Anthony, and Oscar H. Gandy. 2002. All that glitters is not gold: digging beneath the surface of data mining.

and weblining.[119] It is important to note that these are not new and unique market tactics that have emerged due to the use of data mining. However, these arguably unfair[120] strategies rely on the ability of the businesses to classify and sort customers into groups that are useful from the business's perspective. The scale and ease of data processing that data mining allows, as well as the complex and novel patterns and relationships that may emerge enable these business strategies to be employed more easily and more routinely than ever before.

One practice is price discrimination. Danna and Gandy write, "Most economists would agree that price discrimination occurs when the same good or service is sold to different consumers at different prices."[121] Relying on data mining profiles, business can assess which types of customers provide the most profits and which provide the least. Businesses may offer lower prices to customers who match profitable profiles in a bid to keep these customers loyal. Customers who do not match the profitable profiles are not offered special deals.

Price discrimination becomes "customer firing" when businesses try to quietly persuade less profitable customers to leave by raising the costs. Business policies that rely on data mining analysis aim to "fire" the customers who match the least profitable profiles. This tactic is commonly used in the financial service industry where banks often

---

[119] Danna, Anthony, and Oscar H. Gandy. 2002. All that glitters is not gold: digging beneath the surface of data mining.

[120] I have written "arguably unfair" here because some may argue that these practices have a legitimate role in our market economy. This is not the place to engage in such a debate and I proceed as if many would find the increase of such practices disturbing.

[121] Danna, Anthony, and Oscar H. Gandy. 2002. All that glitters is not gold: digging beneath the surface of data mining: p.379.

increase the fees of certain clients in the hopes that the clients will choose to bank elsewhere.[122]

"Weblining" or marketing discrimination is another business strategy made more easy and accessible by data mining technology. "Weblining" is the term for the web-based business form of redlining – when services and business are not offered to potential customers who match certain profiles.[123] For example: a customer who enters some personal information in order to log into a computer company's website may match an unfavorable profile. He may then be denied access to much of their product. If the company has a good site, the customer might not even be aware that he has been excluded from viewing some of the products and from making certain purchases.

The fact that data mining has increased the ease and scale of these questionable business strategies represents an ethical concern. These practices do not give all customers an equal standing in the market. They do not conform to our notion of "fair play". Of course this leads to the question of what is "fair" in the business sector and whether social duties such as fairness and accountability apply. At this point the claim might be made that an increase in these business tactics is not a matter of concern when we consider that they have always had a place in the market. After all, is it not just more of the same? Some of these points might be best addressed elsewhere, for instance within a business ethics context. In this thesis, I will focus on the collective implications of these questionable business practices, and how data mining further complicates these issues.

---

[122] *Ibid.*, p. 381.

[123] *Ibid.*

One reason that the ethical analysis of data mining is complex is because data mining statistics are often used to justify these market tactics. Data mining technology therefore not only enables the questionable business practices but may also provide the rationalization for using them. When I presented a portion of this thesis to a second year Philosophy of Business Ethics class, many of the students were sympathetic to business decisions that may look unfair but are grounded in statistical evidence. For instance, imagine a cell phone company that decides to deny business to anyone who lives in area X. Suppose data mining analysis revealed that seventy percent of customers in area X routinely fail to pay their cell phone bills. The cell phone company concludes that most of their customers in area X do not provide a good return on investment. They decide that it would be too costly to thoroughly check the credit history of potential customers in area X since most of them would probably be denied phone contracts anyway. As a result the company makes it a policy to deny business to anyone who lives in this area. Along comes Sally, who is a longtime resident of area X and who has an excellent credit history. She tries to apply for a cell phone but is not even allowed an application from this company just because of where she lives. This policy to deny cell phone contracts to residents in area X may look unfair to Sally but it may also appear to make good business sense. If it is true that there are solid statistical reasons to support business policies that do not give all potential customers an equal standing in the market, then these policies may look rational and justifiable in our market economy. After all, businesses are in the business of making money.[124]

---

[124] One could still argue that these market practices are unethical in-and-of themselves and that is why I have separated the concerns out as I have. For the purposes of this thesis, I must restrict myself to critiquing the social consequences attributed specifically to data mining technology and I cannot debate the more general ethics of market policies.

However the claim that the statistical evidence provides justification for certain business practices leads to epistemological concerns because the personal data that is collected and mined may be far from complete and accurate. There are many sources of error when collecting and preparing data.[125] There is also no way to separate the good data from the bad or to fill in missing data.[126] Customers may lie about their information; they may provide some false and some true information; or they may not provide all of the information asked for (such as leaving login fields blank). Certain parts of a person's information may change, as, for example, when they move. If the databank is not updated, then this old information might be incorrectly linked to new information.[127] As well, cookies and other web-identifiers do not necessarily identify persons when they collect online information. If computers are shared or someone uses someone else's login password then collected information from one person might be incorrectly linked to another.[128] All of these sources of error are very hard to account for when information is processed. Data mining analysis can therefore lead to the creation of consumer profiles that are misleading and misrepresent groups of people. Inaccurate data could cause improper sorting when a customer is assigned a profile, and might result in the creation of extremely skewed profiles into which all customers are then sorted.

In the short term, this might create negative consequences for the unwary customer. If someone is assigned to the wrong profile, she may be shown products that

---

[125] In the second step of the KDD process the data is rearranged and ordered and sometimes aggregated, such as by postal code, which is known as "preparing" the data for analysis.

[126] Danna, Anthony, and Oscar H. Gandy. 2002. All that glitters is not gold: digging beneath the surface of data mining: p. 378.

[127] For example my Safeway Club card is still linked to my Calgary address. Any information about my Victoria purchases might be linked to this old address, creating an inaccurate picture.

[128] van Wel, Lita, and Lambèr Royyakkers. 2004. Ethical issues in web data mining: p. 132.

do not interest her or she may be denied incentives or other offers she might have received if they had been properly sorted. For example suppose that John wants a cell phone from the cell phone company that I previously mentioned. John lives in region Q but for some reason he is matched to a profile of a region X resident. Since the cell phone company is not doing business with the residents of region X, John is denied an application for a cell phone without any further investigation into his personal information or credit history.

In the long term, business practices based on faulty profiles may wrongly target certain groups of people. Perhaps most of the customers in region X are good credit risks and the profile misrepresents them. All the residents of this region are the victims of a flawed business policy. As I will explain in the section on the larger social costs, decisions that are based on skewed profiles may lead to the creation of vulnerable groups of people or the further devaluing of those already marginalized.

At this point, it may look as though the major problems associated with data mining are due to inaccuracies in the consumer profiles. If we accept that business policies are justifiable when they are based on sound market research (even if they are not "fair") then it looks as though, once data mining technology is "fixed" and the customer profiles accurately represent the market, there will be no more ethical concerns associated with data mining.[129,130] However serious societal harm can result from categorization and customer profiling even if the technology is "fixed". I will explain these consequences in

---

[129] This assumes that the increase in these questionable business policies is not cause for alarm in and of itself. As I have stated, this is not an issue I can address in this thesis.

[130] After all, large error margins do not provide optimal market analysis for businesses. Data mining technology is constantly advancing and skewed profiles may come to be a thing of the past. For now, the epistemological concerns are still relevant in this ethical analysis.

more detail now to show not only how the questionable business practices and the epistemological concerns that I have outlined compound the larger ethical issues, but also that the long term social implications associated with data mining represent a serious independent concern.

Serious social problems are associated with customer profiling because of the practice of grouping and discrimination. Oscar Gandy Jr. points out that categorization leads to consequences.[131] He writes,

> "A profile is primarily a list of categories that have been determined to be relevant to some administrative decision that must be made by an organization with regard to an individual, a group, or a class.... The fundamental purpose of a profile is the assignment of an individual into a class or category that represents a decision. This is a process of identification and consequences."[132]

The social consequences of such profiling can be great and it is worth going into them in more detail and with more examples than I have previous discussed.

When a decision is made as to which groups will be included in special offers and promotions then the corresponding decision to exclude certain groups is also made. In some cases, this may look benign and might arguably be preferred by business and consumers when dealing with mass advertising campaigns. However, as we have seen, it may be the case that groups of people are denied an equal opportunity to participate in various promotions or denied even the opportunity to do business with an organization. Instead of being denied a cell phone contract or a bank loan because of certain financial criteria, people who are grouped by statistical criteria may be denied even the *offer* to do business with the organization. Profiles may be based on a combination of particular

---

[131] Gandy, Oscar H. Jr. 2000. Exploring identity and identification in cyberspace.

[132] *Ibid.*, p. 1099.

markers such as income, race, age, gender, and postal code. Other more obscure measures may also be significant such as the type of car you drive, the number of children you have, whether you eat out more than twice a week, whether you have pets, or other unfathomable criteria that make up novel data mining patterns.

Some actual and hypothetical scenarios might help to elucidate the problems of data mining and group profiling. A good hypothetical example is offered by Tavani who tells the story of Lee, a junior executive who applies for a car loan to purchase a BMW.[133] Lee fills out the necessary paper work including relevant salary and loan history information – he is currently making $90 000 per year in a new job as a marketing executive and is repaying a $15 000 loan used to finance a family vacation. The bank's request for this information is appropriate as it is necessary for their assessment of whether or not Lee should be granted the loan. Lee provided the information for the purpose of the bank's credit assessment so that he can secure the loan. The bank takes Lee's information, mines it with data from the bank's warehouse and discovers the following pattern:

> "executives earning more than $70,000 but less than $120,000 annually, and who purchase luxury cars (such as BMWs), and who take expensive vacations, often go into business for themselves within five years of employment. A separate pattern-matching program reveals that the majority of marketing entrepreneurs who go into business for themselves declare bankruptcy within one year of starting their own businesses. All of a sudden, Lee is a member of a group that neither he nor possibly even the officers at the bank had ever known to exist, viz., the group of marketing executives likely to start a business and declare bankruptcy within a year of starting such a business. With this 'new information' about Lee, the bank determines that Lee, and people that fit into Lee's group are long-term credit risks."[134]

---

[133] Tavani, H.T. 1999. Informational privacy, data mining, and the internet. *Ethics and Information Technology*, **1**: 137-145: p. 141.

[134] *Ibid.*, p. 141.

Thus Lee freely gave the bank information about him and authorized the use of this information to access his credit standing. However, the bank did not rely on the explicit information that Lee provided but rather on implicit patterns in their data base. Lee was not refused his loan due to his personal information but rather through the use of implicit patterns from the data of people similar to Lee in certain respects but vastly different from Lee in other important respects.[135] The profile was used by the business as if it was more meaningful than the individual data. Lee was not treated as an individual; he was treated as a member of a category. He was not given equal consideration as a person; he is only given equally discriminatory consideration with those who also match this profile. As I will discuss later, categorization not only affects individuals, it can also have serious societal implications.

Roland Pierik describes a real world example of some consequences of data mining and group profiling.[136] He states that some Dutch cell phone companies refuse to provide various people with cell phone contracts based on their postal codes. Those deemed to be living in neighborhoods with poor average payment records are denied services regardless of their personal payment behavior.[137]

Danna and Gandy make note of a woman who was denied a gasoline credit card because the profile included her zip code.[138] She was a white woman who lived in a predominantly black neighbourhood – a neighbourhood that received the lowest of five

---

[135] *Ibid.*, p. 141.

[136] Pierik, Roland. 2001. Group profiles, equality, and the power of numbers. In Anton Vedder (ed.). *Ethics and the Internet*, Antwerpen-Groningen, Oxford: Intersentia.

[137] *Ibid.*, p. 117.

[138] Danna, Anthony, and Oscar H. Gandy. 2002. All that glitters is not gold: digging beneath the surface of data mining: p. 382.

ratings on the gasoline company's rating system. The woman filed a civil rights case; however, she lost because she was not able to prove that the decision to deny her the credit card was based on racial discrimination against the area in which she lived.

These examples show that group profiling can result in business policies that do not view customers as individuals. When consumers are grouped and classified, their individual characteristics that set them apart from the profile may no longer matter. This is analogous to the way that statistical results are often misinterpreted. It is a common mistake when interpreting statistics to assume that any one individual exhibits the characteristics of the group. In economics this is known as the ecological fallacy.[139] An example of this would be to assume that any one person in a certain neighborhood makes around $30 000 if the average wage in the area is $30 000. Of course the actual wage of any one person could be quite far from the average. (For instance one person could make $10 000 while another makes $50 000, making an average wage of $30 000). In terms of data mining, the ecological fallacy occurs when it is assumed that any member of a consumer profile exhibits all the characteristics of that group. Remember that the profiles used by businesses are mostly non-distributive, which means that not all of the characteristics are true for every member of the group. There is no guarantee that a person assigned to a group will behave as the profile predicts they will. For example, a business might think that if region X is a place where seventy percent of cell phone customers have been poor credit risks in the past, then Sally is a poor credit risk *because* she lives in region X. As such, the business might assume that none of Sally's

---

[139] Robinson, W.S. 1950. Ecological correlations and the behavior of individuals. *American Sociological Review,* **15:** 351–357.

other characteristics matter in their assessment of her. Since she lives in region X, then she is a poor credit risk.

One may counter this notion of the ecological fallacy with the claim that these businesses are not misinterpreting statistics; it is just that these policies are made on good business sense. Even if Sally is a good credit risk, the cell phone company may claim that it does not pay for them to check everyone's credit history in an area where a high percentage of customers do not provide a good return on investment. While the ecological fallacy is not necessarily occurring in all business decisions, it may still pertain when other businesses and individuals interpret these policies. Other companies may assume that the area is a poor place to do business, and people who live outside this area may associate all the people in this region with the negative profile.

More importantly, serious problems can result whenever customers are not treated as individuals but as members of certain categories. Custers writes,

> "When you are a member of a group, others suppose that you have the characteristics on which the group profile is based. When you do not have these characteristics, it is up to you to prove this. This is a guilty-until-proven-innocent system, which should be avoided."[140]

As in the case of the woman in the U.S. who was denied the gasoline credit card[141], there is often no recourse for those who are disadvantaged by a business policy. Once you are assumed to belong to a group there may not be much you can do to show that you do not belong to that group. If this means that you are prevented from an equal place in the market then the immediate negative consequences for you could turn into long lasting social consequences that I will discuss in more detail later.

---

[140] Custers, Bart. 2001. Data mining and group profiling on the internet: p. 95.

[141] Danna, Anthony, and Oscar H. Gandy. 2002. All that glitters is not gold: digging beneath the surface of data mining: p. 382.

Information technology and our networked society may further complicate the consequences of categorization. In a lecture I gave on data mining a student told me that he had been mistakenly blacklisted by his bank's credit rating system. The bank acknowledged their error and they told him that they could fix it on their end; however they could not alter the networked system, which had now labeled him as a bad credit risk. He was told that he would have to wait the eight years that it took for a person's credit status to change before other creditors would overlook his rating. Though this example may not be related to data mining (and may not be entirely accurate), it works as a thought experiment to highlight that networked systems might make it even more difficult to prove to all those involved that you do not match a certain profile. As well, it demonstrates that ubiquitous information sharing may make the negative social consequences of categorization even more severe.

We have seen that data mining can lead to problems within the market. Customer profiling and categorization can also cause severe and lasting repercussions for the whole of society. The reasons for this are due in part to the way that humans necessarily order and classify the environment. Peirik writes, "The purpose of classification is to reduce the infinity of possible differences in one's environment to workable proportions, while maintaining relevant discriminations between classes."[142] Classification is a necessary and often invisible way of organizing the world so that we can move through life's numerous complexities. Without grouping objects into various categories, the world would be seen as an incomprehensible mass of stimuli. Categories are so fundamental to everyday life that it is not as though we think about them or act upon them, rather, we act

[142] Pierik, Roland. 2001. Group profiles, equality, and the power of numbers: p. 107.

within them.[143] These classifications are not neutral groupings, they are judgments that we make about the various stimuli we encounter. In other words, "to cognize is to categorize"[144]:

> "We organisms are sensorimotor systems. The things in the world come in contact with our sensory surfaces, and we interact with them based on what that sensorimotor contact "affords". All of our categories consist in ways we behave differently toward different kinds of things -- things we do or don't eat, mate-with, or flee-from, or the things that we describe, through our language, as prime numbers, affordances, absolute discriminables, or truths. That is all that cognition is for, and about."[145]

Categorization relates to data mining because the consumer profiles are groupings based on statistical analysis.

> "Statistics create new phenomena in society, namely social categories. These categories – e.g., yuppies or unmarried mothers – and the characteristics and labels associated with these categories cling to a person's memories. Data is translated into 'information,' and this is interpreted as 'knowledge.' These categories get fixed in public knowledge, even if in the course of time the information on which they were based prove to be outdated or simply untrue."[146]

The group profiles produced through data mining may take on a life of their own entirely apart from the data bases where they originated, contributing to stereotyping and the devaluing of various groups beyond the marketplace.[147] For example, if it is known that a company has stopped doing business in a certain area, people might come to think of this area and the people in it as undesirable. If the area is made up of people who are

---

[143] Pierik, Roland. 2001. Group profiles, equality, and the power of numbers: p. 108.

[144] Harnad, Stevan. 2005. To cognize is to categorize: Cognition is categorization. In Henri Cohen and Claire Lefebvre (eds.), *Handbook of Categorization in Cognitive Science*, Elsevier, Amsterdam: p. 1.

[145] *Ibid.*, p. 1.

[146] Pierik, Roland. 2001. Group profiles, equality, and the power of numbers: p. 114.

[147] *Ibid.*, p. 114.

already disadvantaged, then these business policies might be seen as evidence that endorses negative stereotypes. As we have already seen, the woman in the U.S. who was denied the gasoline credit card believed that this decision was based on the racial stereotype of her largely ethnic neighborhood.[148] She filed a human rights claim but did not win her case because she could not prove the business decision was racially motivated. Business decisions like this one are detrimental to groups that are already marginalized and to all the people who fit into "unprofitable" or otherwise "unfavorable" profiles. The negative consequences may be compounded if other businesses blindly adopt similar policies.

These profiles and stereotypes are also very hard to get rid of. This is because of the necessary role that classification plays in our understanding of the world. Another reason is that data mining analyses are presented as statistics and facts. The profiles have, "a touch of rigor, universality, impersonality, impartiality, objectivity, neutrality, and maybe even truth. This suggestion is strengthened by the fact that the results…are presented in clear numbers, 'undeniable facts'."[149] Statistics and scientific claims are often accepted as "privileged knowledge" within our society.[150] When findings are interpreted outside of the field where they originated, then the validity and merit of the claims may not be assessed or presented. In other words, scientific and statistical claims

---

[148] Danna, Anthony, and Oscar H. Gandy. 2002. All that glitters is not gold: digging beneath the surface of data mining: p. 382.

[149] Pierik, Roland. 2001. Group profiles, equality, and the power of numbers: p. 114.

[150] This is sometimes referred to as "scientism" or the "scientistic fallacy" (reference to follow). It is usually discussed within a theological critique of the claim that only science can provide knowledge of reality. It is relevant here for the following reason: "At the heart, the scientistic fallacy can be understood as a kind of category mistake that occurs when claims from academic discipline are extended beyond their proper bounds" (Peterson, 755). Peterson, Gregory R. 2003. Demarcation and the scientistic fallacy. *Zygon: Journal of Religion and Science*, **38**(4): 751-761.

may be accepted as "fact" when they were not originally presented as such. Thus denying certain groups of people equal standing in the market could lead to serious long term consequences outside the market. The use of statistical justification may also create a way for businesses to knowingly hide various unaccepted practices such as blatant discrimination. If companies only want to do business with certain types of people, they may seek out profiles that "justify" denying service to the customers they see as less desirable. In these ways profiling and categorization could cause the de-valuing of certain groups of people, the perpetuation of stereotypes, and the further marginalization of those already disadvantaged.

These social consequences of data mining are not necessarily unique. They can occur in all matching and profiling practices.[151] However, it is the scale of data that can be processed through data mining that makes the resulting social consequences that much more serious. As well, data mining techniques are novel in that they can uncover implicit information and connections. Spinello and Tavani write, "Relatively new…are the ever-growing possibilities of discovering hitherto unnoticed relationships between characteristics and features of persons created by [data mining]."[152] The problem is that data mining can create new vulnerable groups of people and it can also further perpetuate discrimination and subjugation of groups that are already marginalized.

A final point of critique in the ethical analysis of data mining is based on the claim that consumers are often not properly informed before they consent to providing personal information. In our society, personal information has become a commodity that many

---

[151] Spinello, Richard A., and Herman T. Tavani (eds.). 2004. *Readings in Cyberethics*. Jones and Bartlett: p. 4.

[152] *Ibid.*, p. 4.

businesses have cashed in on, to the point that they may be taking advantage of customers. Privacy policies are often ambiguous and open to interpretation. For example many companies state that they will not sell or share your personal information with "third parties". This may mean that they *will* sell or trade it to shareholders or businesses that have ties to the company that collected the information. These secondary businesses may feel that they have no privacy obligations with regards to your information; they may turn around and sell it to someone else. If this is the case, then protection against "third parties" has not really kept your personal information private.

It is also questionable whether all information collecting methods are apparent and agreed to by customers. Information may be collected in extremely subtle ways such as a store clerk coding in your gender and approximate age along with your purchase.

In addition, special offers or incentives to provide information may raise questions of coercion. It is a common business practice to withhold goods or services until personal information is provided. This is the case in website logins where information must be supplied in order to access the site. It is also true of many grocery store client cards where customers are not eligible for certain discounts unless they use cards that enable their personal information to be connected to their grocery purchases. It may be that the incentives and discounts offered by businesses in exchange for personal information constitute undue enticement. These questions of coercion, the ambiguous privacy policies, and subtle information collection methods all raise concerns that data mining does not respect the informed consent of the consumer.

In conclusion, there are serious long-term social consequences associated with the use of data mining. This technology enables businesses to sort consumers into various

profiles which can lead to de-individualization. Consumers who match the profiles deemed to be less profitable to a business may be denied incentives or services, or may be pressured to quietly end business with the organization. In the short term, this may lead to negative consequences for the unassuming consumer. In the long term, this practice may cause serious societal harm if the profiles pervade our social consciousness and promote discrimination and de-valuing of certain groups of people.

It may be apparent from this ethical analysis why some critics argue that the best ways to articulate the problems associated with data mining are in terms of discrimination, human rights violations, and a lack of informed consent. Businesses policies that prevent certain groups of people from participating fully in the market are discriminatory. When businesses use data mining analyses and statistics to cover up or perpetuate prejudicial practices then human rights are being violated. If customers are coerced into providing information or are misled as to how their information is collected then these customers may not be freely consenting to provide their information. It is easy to see that these points of critique represent valid concerns, yet it may not be clear what the role of privacy is in this analysis. In the next chapter I will explore the contentious connection between privacy, data mining, and the associated social consequences by returning to the conception of privacy that I outlined in Chapter 1 – privacy as limited access to the person.

Chapter 4:

Understanding the Relationship between Privacy and Data Mining

In the last chapter, I explained some of the negative consequences and social implications that are associated with certain uses of data mining. Data mining technology has enabled more consumers to be labeled and classified than ever before. This has been very advantageous for businesses; however, categorization leads to consequences. Profiling practices can promote discrimination and the de-valuing of certain groups of people. When market policies target areas and people who are already disadvantaged, data mining analysis may appear to endorse stereotypes and may perpetuate prejudice.

These serious social outcomes, and the fact that it is hard to see how privacy is a factor in this debate, have led some critics to claim that privacy violations are not the major concern. They argue that the focus should shift away from privacy to other ethical frameworks such as discrimination, human rights violations and informed consent as these might provide more accurate and useful views of the problems, and more compelling positions from which to work for social change.

In this chapter, I will explore the contentious connection between privacy and data mining by returning to the privacy definition I outlined in Chapter 1 – privacy as "limited access to the person". I will examine what this privacy conception can say about the relationship between data mining and privacy and also what it can say about our current state of privacy in this information age.

Return to Privacy as "Limited Access to the Person"

In Chapter 1, I stated that one way to explore connections between privacy and technology is to see whether accepted definitions of privacy adequately address the ethical issues associated with new technologies. In other words: Can the concerns raised by the technology be properly articulated as privacy concerns within the context of the accepted definition of privacy?

When privacy definitions do not provide a good context for articulating the ethical issues associated with a new technology then it is important to determine the reasons for this. It could be that the technology is raising concerns that may appear to be privacy issues but may not be privacy issues at all. If this is the case, then grouping these concerns under a "privacy" heading causes conceptual confusion that needs to be addressed.

Alternatively, it could mean that the new technology is eroding our privacy. If this is the case, then the accepted definition of privacy might be lacking; perhaps this traditional way of conceiving privacy is now inadequate in our technological society. It might be that our privacy conceptions are not robust enough to protect our privacy in this information age.

In this chapter, I will return to the privacy definition I outlined in Chapter 1 – privacy as "limited access to the person" – to explore whether or not there is a connection between privacy and data mining. Recall that this conception of privacy is endorsed by Ferdinand Schoeman[153], developed and elaborated by Ruth Gavison[154], and amended

---

[153] Schoeman, Ferdinand. 1984. *Privacy: Philosophical dimensions.*; *and:* Schoeman, Ferdinand (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology.*

[154] Gavison, Ruth. 1980. Privacy and the limits of the law.

from a feminist perspective by Anita Allen[155]. To briefly reiterate, Ruth Gavison states that secrecy, anonymity and solitude are all distinct yet inter-related components of privacy conceived of as limited access to the person.[156] Secrecy involves limited access to information about a person.[157] Anonymity involves limited access to a person through limited attention paid to that person. Solitude involves limited physical access to a person. Privacy can be lost through the loss of one of these aspects, or two or three in combination. I will at look each of these irreducible elements of privacy to see how they relate to the ethical critique of data mining. After this assessment, I will explore Allen's amendment to Gavison's conception of privacy to see if Allen's addition is helpful to the discussion at hand.

The first component of privacy as limited access to the person is secrecy – secrecy involves limited access to information about a person. This element denotes that privacy is related to the amount of information known about a person.[158] The information collected for data mining purposes does seem to be somewhat excessive. Does data mining therefore violate privacy with respect to secrecy? Firstly, it does not appear that the *type* of information collected for data mining purposes violates privacy. Data mining

---

[155] Allen, Anita L. 1984. *Women and their privacy: What's at stake?*

[156] Gavison, Ruth. 1980. Privacy and the limits of the law: p. 351.

[157] Remember that the term "secrecy" is not being used in the way that we might typically think. Gavison is using the term to mean something like the contemporary notion of "informational privacy".

[158] Gavison discusses that it is the "amount" of information that is relevant. She accepts that this raises epistemological questions over how one quantifies "excessive" information but she states that most cases are quite straightforward and thus this seems to be more of a theoretical problem than a normative one. Gavison, Ruth. 1980. Privacy and the limits of the law: p. 351.

relies on largely public information that is not personal in nature. Based on Gavison's discussion, this sort of information does not raise privacy issues in and of itself.[159]

The claim could be made that it is not the type of information; rather the *amount* that is collected is "excessive" and therefore constitutes a privacy violation. Gavison does state that large Government databanks involve a loss of privacy.[160] However, Gavison also states that privacy is lost "as information *about an individual* becomes known" (emphasis mine).[161] Government banks contain information that identifies you and thus it is information about you. The fact that the information used for data mining purposes is firstly public and secondly de-identified,[162] makes it hard to show how this collection and use constitutes gaining excessive information "*about a person*". If my age, gender, and postal code are linked to a purchase I make and this information is then de-identified and mined, all that would be known is that a woman in approximately this age

---

[159] Gavison discusses an example where a loss of privacy occurred when the contents of a person's love letters were published without his permission, and an example where a loss of privacy occurred when an employer asked a series of inappropriate questions of his employee. Unlike data mining, these examples involve information that is sensitive or "private" in nature. Gavison, Ruth. 1980. Privacy and the limits of the law: p. 351.

[160] *Ibid.*, p. 351.

[161] *Ibid.*, p. 351.

[162] It is important to note that the true "de-identification" of data may be becoming a misnomer as technology advances and "re-identification" is now possible. Through the use of algorithms and other techniques, data that is devoid of all *explicit* identifiers (such as name, address, and telephone number) can be used to re-identify people with an alarming success rate (*see for example*: Sweeney, Latanya. 2000. *Uniqueness of Simple Demographics in the U.S. Population*, LIDAPWP4. Carnegie Mellon University, Laboratory for International Data Privacy, Pittsburgh, PA. Forthcoming book entitled, *The Identifiability of Data*.) As such, some authors, notably Latanya Sweeney, (*see for example*: Sweeney, Latanya. 2002. *k*-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems,* **10**(5): 557-570.), argue that certain "de-identified" data should not actually be treated as if the data subjects are truly anonymous (i.e. the data should not be sold, traded, or aggregated without thought to the way that information can be linked and people can be identified). Re-identification is especially troubling when we consider that hospital records and other extremely sensitive information can be re-identified using largely public information that is provided in different, seemingly unrelated settings. As technology advances, we will need to monitor how 'anonymous' de-identified data truly is and what the ethical implications are in our highly networked society.

range, who lives in this area, bought this. Unlike a large Government database, it would be very hard for anyone to look into the company's data warehouse and know that information was about me. In this way, no one has really gained excessive information *about me* per se. For these reasons, privacy does not seem to be lost due to the amount or type of information collected for data mining purposes – the component of secrecy does not capture the ethical problems associated with data mining.

If it is neither the type nor the amount of information that is the issue, then the problems may lie in the way that the information was collected. This concern is best addressed with respect to "anonymity" – the second component of Gavison's definition of privacy. Anonymity involves limited access to a person through limited attention paid to that person. Gavison writes,

> "An individual always loses privacy when he becomes the subject of attention. This will be true whether the attention is conscious and purposeful, or inadvertent."[163]

Is excessive attention paid to an individual when his data is collected? It looks as though this may be the case. The average person would probably take issue with someone following him around the store writing down every detail they could – what was put in the cart, the sex and approximate age of the shopper, whether there were children present or not, a rough guess at household income based on general appearance and clothing style of the shopper. This does seem like harassment and a violation of our anonymity; and Gavison states that this is how losses of privacy occur – when attention is paid to someone in a direct way such as following him, listening to him, observing him, etc.[164]

---

[163] *Ibid.*, p. 353.

[164] *Ibid.*, p. 353.

However, what if the "spy" asked if she could do this because she works at the store and it helps with customer service? What if she said they could do it non-intrusively so that the collection of information would be hardly noticeable? The claim is that this is indeed what happens: either explicitly or implicitly consent is obtained. One has to sign up for a store client card or credit card, and privacy policies abound detailing what information is collected and how it is collected. If there is consent, it would be hard to argue that there is a violation. Furthermore, Gavison states that when attention is paid to individuals in less direct ways, such as discussing, imagining, and thinking about an individual, then these activities violate privacy only to the extent that they lead to greater interest and more attention paid *to the individual.*[165] Since information is collected for data mining purposes in very unobtrusive ways, it is at most, indirectly related to anonymity. Since the information is then de-identified, it does not lead to greater attention paid to the individuals from whom it was collected. Thus privacy does not seem to be lost due to excessive attention paid to the person and the component of anonymity does not seem to capture the ethical concerns associated with data mining.

What about the claim that the information is not always "freely" given? As we saw, there were worries over ambiguous privacy policies, the subtle or furtive ways that information is collected, and the coercive tactics used to obtain information. The way that information is collected and used for data mining purposes might raise concerns over informed consent. Though this issue represents a valid concern, it would be hard to formulate it into a strong critique in terms of privacy as "limited access to the person". I have explained that the complaint could not be formulated in terms of excessive

---

[165] *Ibid.*, p. 354.

information or excessive attention. If the claim is that people are being taken advantage of then it looks as though Gavison would not see this as a privacy issue. In the section where Gavison explains what privacy is not, she states that "commercial exploitation" does not constitute a privacy violation.[166] It is because the information that is collected and used for data mining purposes is mostly public and de-identified before use, that the concerns relating to informed consent of the customer could not be formulated as involving losses of privacy, in Gavison's terms.

The final component of Gavison's definition of privacy involves solitude or limited physical access to a person. Data mining does not seem to violate one's solitude as no excessive physical access is involved.

Does this mean that Gavison's definition of privacy does not address the ethical issues raised by data mining? Recall that Gavison's arguments outlining privacy as limited access to the person are convincing because of the way her definition of privacy functions. She argues that, so conceived, privacy allows for autonomy; the promotion of mental health, human relations, and liberty of action; and the freedom from ridicule, censure, and physical access.[167] Limited access to a person based on the dimensions of secrecy, anonymity, and solitude allows an individual to create an independent identity apart from possible hostile reactions of others. Such an identity is necessary in order for an individual to become a fully realized person. Thus Gavison writes,

> "We desire a society in which individuals can grow, maintain their mental health and autonomy, create and maintain human relations, and lead meaningful lives. [My] analysis…suggests that some privacy may

---

[166] *Ibid.*, p. 356.

[167] *Ibid.*

> therefore indicate the existence of and contribute to a more pluralistic, tolerant society."[168]

Yet certain applications of data mining do not promote the full realization of autonomous people. Data mining technology enables businesses to sort consumers into profiles and treat them, not as individuals but as members of a category. It is acceptable that businesses are interested in you for what you can do for them, for that is the nature of the game. The problem is that groups of people have identities imposed upon them and are treated according to how those identities are *predicted* to act. It seems clear that society does not become more egalitarian as a result of profiling. It also seems clear that data mining does not promote autonomy when there may be no option whether to enter into certain business transactions or not (for instance when the application for a credit card is denied based on the location of your house) or possibly to even know that such business transactions exist (for instance, when a web based computer company limits the products available to you based on the profile you were sorted into).

Of course the claim that privacy is necessary for an egalitarian society does not mean that if society is lacking egalitarianism then its problems necessarily stem from a lack of privacy. Thus Gavison's intuitively satisfying definition of privacy that accounts for the creation of an independent identity and the promotion of autonomy does not mean that if data mining prevents the realization of this independent identity then data mining must be affecting privacy. It is, however, important to note this connection as we proceed to uncover whether privacy issues underlie the ethical concerns associated with data mining.

---

[168] *Ibid.*, p. 369.

In Chapter 1, I also discussed Anita Allen's feminist critique of Gavison's account of privacy.[169] Allen accepts Gavison's argument that privacy involves limited access through the dimensions of secrecy, anonymity and solitude. However, Allen contends that Gavison's definition, as with many other privacy definitions, lacks certain requirements that are necessary to protect the privacy of women. Allen argues that what women need to secure privacy is freedom of choice. She states that without the ability to choose with respect to sex, childbearing, and marriage, women will not be able to secure privacy.[170] Thus Allen argues that a definition of privacy that promotes the dimensions of secrecy, anonymity, solitude, and choice are necessary for the protection of women's privacy interests and for the promotion of liberal ideals.

In relation to data mining, your gender, whether or not you have children, and whether or not you are married may all be part of the construction of a profile or the assignment to one, but the notion that women require special privacy protection does not capture the heart of the ethical critique.[171] One company may offer women with children a better deal than single men. Again, the combinations, parameters, patterns, and profiles are largely implicit and unknown before analysis begins. After processing the data, it is the marketing decisions that determine what impact the profiles will have on consumers and society in general.

---

[169] Allen, Anita L. 1984. Women and their privacy: What's at stake?

[170] *Ibid.*, p. 244.

[171] Any component used in the construction of, and assignment to, a profile may not necessarily be discriminating. As novel patterns are uncovered within the data, businesses may offer incentives to different groups of customers at different times. I will discuss later however, that the context and the relevance of information is different for different groups of people. For example, personal information often carries more consequences for the marginalized then it does for the dominant classes.

Nevertheless, Allen's analysis does highlight an interesting idea. While I do not think that data mining highlights that women, per se, need special privacy protection, Allen's notion that members of society might be disadvantaged more than others with respect to privacy, is an interesting one. As we saw, certain data mining applications disadvantage certain groups of people more than others. Categorization leads to consequences and profiling practices can promote discrimination and the de-valuing of people. When business policies target areas, and people who are already disadvantaged, data mining analysis may appear to endorse stereotypes and perpetuate prejudices. The point is that the collection and use of information, even largely public, non-identifying information, may negatively impact certain groups of people more than others. Does this mean that data mining is violating the privacy of some groups of people more than others? Should Gavison's conception of privacy as limited access to the person be further amended to include special privacy considerations for certain groups of people?

It is still hard to see how the disadvantaging of certain groups of people can be viewed as a matter of privacy. If my information is used to create a business policy that disadvantages other people, how is my privacy affected, or theirs for that matter? The people who have been disadvantaged might not even have contributed information to the policy. Sally's information may not have contributed to the cell phone company's decision to deny contracts to people who live in region X, yet it is Sally who is prevented from getting a cell phone. Furthermore, customer profiles are based on large amounts of aggregated information. No one person could easily show how their information contributed to a policy, so how would they be able to claim that their privacy had been affected?

Based on this analysis, it looks as though Gavison's conception of privacy is not a good framework to express the problems associated with data mining. However, recall from Chapter 1, when privacy definitions do not provide a good context for articulating the ethical issues associated with a new technology, it is important to determine why. It may be that the technology is raising different issues that cannot and should not be grouped under a privacy heading; some critics claim that this is what is happening with data mining. However, it could also be that our current conceptions of privacy might be lacking with respect to the new technology. I intend to show that the philosophical difficulties we are encountering are due to the fact that traditional ways of conceiving privacy cannot fully secure privacy in our technological society.

It is important to note that I am not attempting to reject traditional definitions of privacy; I accept that these ways of viewing privacy are still important and very valuable for the civil liberties they provide to the individual. Rather, my argument is that information technologies like data mining highlight that our common privacy conceptions focus on the benefits that privacy provides to the individual and see the benefits to society as the by-product of this more important relationship. As such, the inherent social value of privacy has been largely ignored in our laws and policies. Overlooking the social significance of privacy has led to problems in securing privacy in our networked economy. This has created an unequal distribution of the burdens and benefits of information technologies such as data mining. Therefore, my focus is not to reject the traditional ways of conceiving privacy, nor am I proposing to redefine privacy; rather I mean to draw attention to the way that privacy is an important social value that needs to be acknowledged in our laws and policies in order to provide equitable privacy protection

and an equitable distribution of the benefits and burdens of information technologies such as data mining.

The nature of my claim will become more apparent throughout this chapter and in Chapter 5, where I will clarify the connections between privacy, data mining and the associated negative social consequences. Before I discuss the foundations for this argument, I will briefly describe what I mean by the "social value" of privacy. Regan writes, "Privacy is rapidly becoming a *collective value* in that technology and market forces are making it hard for any one person to have privacy without all persons having a similar minimum level of privacy."[172] In this way, "If one individual or a group of individuals waives privacy rights, the level of privacy for all individuals decreases because the value of privacy decreases."[173] Discussing privacy in a similar context, Bennett and Raab write,

> "it is the *collective*, indivisible value of privacy which is particularly important, and which is perhaps especially vulnerable to attack by inequalities. As a collective value, privacy cannot easily be provided to one person without its being enjoyed by others. Rather like street lighting, to provide for one is to provide for all."[174]

The shift from the focus solely on the individual to one that recognizes the collective or social nature of privacy is conceptually challenging because of the way we have come to think of privacy in relation to the individual and the individual's role in controlling and being affected by information *about herself.* As Anton Vedder writes, "The one-sidedness that comes with the privacy vocabulary is apt to create an imbalance

---

[172] Regan, Priscilla M. 1995. *Legislating Privacy: Technology, Social Values, and Public Polic:* p. 213.

[173] *Ibid.*, p. 233.

[174] Bennett, Colin J. and Charles D. Raab. 2003. *The Governance of Privacy: Policy Instruments in Global Perspectives*. Hampshire, UK: Ashgate Publishing Limited: p. 41.

in our capacities to perceive, analyze and articulate the multifarious moral aspects of information technology."[175] Many contemporary writers[176] argue that the nature of privacy is not solely grounded in the importance or the protection it provides to the individual but we have come to think of privacy only in these terms because privacy has been so adamantly articulated and protected this way. Bennett and Raab state, "The social value [of privacy] is underpowered and survives precariously unless it can be specifically reinforced by a change in the privacy culture, for it is powerfully challenged by the legacy of the conventional paradigm and by forces that tend to the protection of privacy seen as an individual value, if a value at all."[177]

The challenge is therefore to understand how privacy can be regarded as an important *social value* and not just important because of how it relates to individuals.[178] The conceptual framework for this notion will be further established in Chapter 5 where I will discuss the idea that privacy can be viewed in a similar way to a collective good. Presently, I will explain the *prima facie* reasons for the claim that privacy has a larger social significance that is often overlooked in our laws, policies and philosophical definitions. I rely on two interrelated arguments: The first argument is based on the

---

[175] Vedder, Anton. 1999. KDD: The challenge to individualism: p. 280.

[176] *See for example*: Bennett, Colin J. and Charles D. Raab. 2003. *The Governance of Privacy: Policy Instruments in Global Perspectives*.; Vedder, Anton. 1999. KDD: The challenge to individualism.; Regan, Priscilla M. 1995. *Legislating Privacy: Technology, Social Values, and Public Policy;* French, Martin. 2003. Privacy across cultures: Indigenous self-determination and the politics of information. *Unpublished Thesis*. University of Victoria, Victoria, B.C.

[177] Bennett, Colin J. and Charles D. Raab. 2003. *The Governance of Privacy: Policy Instruments in Global Perspectives*: p. 41.

[178] As we saw in Chapter 1, Gavison and others believe that privacy is important to a free and egalitarian society but, as I will argue later in this chapter and in Chapter 5, this understanding of the importance of privacy is still grounded in the notion that privacy is valuable to individuals as individuals. Thus, notions such as Gavison's do not adequately acknowledge the larger social value of privacy. Recognition of the collective value of privacy is needed if privacy is to be protected in our networked economy.

common sentiment and intuitive connection between privacy, information technologies, and the associated negative social consequences. The second is evidence that our traditional definitions of privacy and the way that privacy is protected in our laws and policies focus on the importance of privacy to the individual. This has prevented the larger social value of privacy from being properly acknowledged which, in turn, has led to the further repression and subjugation of certain marginalized people.

Firstly, even though there appears to be conceptual confusion about how privacy relates to data mining and information technologies, many still think that a connection exists. Technologies such as data mining stir "popular indignation, worry and resentment"[179] and our concerns surrounding these technologies are usually discussed in terms of privacy. We hear such things as, "What is the state of our privacy in the information age?" "Is technology threatening our privacy?" In the quoted passage at the beginning of this thesis Scott McNealy responded to criticism over his company's new networking software with the universal statement, "You have no privacy. Get over it."[180] The way we think and talk about the connection between privacy and information technology may be evidence that we have a normative notion of privacy that is not fully accounted for by definitions of privacy such as Gavison's that focus on privacy as limited access to the person. The way that we think and talk about privacy in connection with information technologies may also be evidence that we have collective concerns about privacy that need to be addressed.

---

[179] Nissenbaum, Helen. 1998. Protecting privacy in an information age: The problem of privacy in public: p. 579.

[180] Scott McNealy made this comment at a 1999 Sun Microsystems news conference, during the unveiling of the company's latest software; a new product with great networking abilities called, Jini. (Widely discussed on the internet and cited in many articles including: Austin, Lisa. 2003. Privacy and the question of technology. *Law and Philosophy,* **22:** 119-166.)

Many surveys and polls have shown that the public is truly concerned that information technologies are a threat to privacy.[181] Yet this sentiment is not echoed in the conception of privacy as "limited access to the person". Still, the fact that there is a common sentiment about the connection between privacy and information technologies will only have normative force if the underlying reasons for this are grounded in a moral basis. Nissenbaum explains this in her discussion on the common sentiment about information technologies and privacy:

> "We may regard public expression as a sign, of something more than preference and mere opinion – more so if it is consistent and fairly widespread – and we must seek a greater understanding of its source…To suggest a moral basis for expression of popular indignation we must show that popular reaction plumbs human needs that are deeper and more universal than "mere" preferences and desires."[182]

I intend to show that this common sentiment about privacy and information technology indicates that privacy has a collective value that is not adequately captured in our notions of privacy and I draw on a critique from the margins to provide support for this claim. Many philosophers believe that a marginalized standpoint can offer valuable insight into the underlying assumptions of society.[183] In this case, valuable social commentary is presented in an analysis of the impact that dominant privacy conceptions

---

[181] See for example: Regan, Priscilla M. 1995. *Legislating Privacy: Technology, Social Values, and Public Policy*.; and Nissenbaum, Helen. 1998. Protecting privacy in an information age: The problem of privacy in public.

[182] Nissenbaum, Helen. 1998. Protecting privacy in an information age: The problem of privacy in public: pp. 580-581.

[183] *See for example:* Lennon, Kathleen. Feminist epistemology. In Ilkka Niiniluoto (ed.). 2004. *Handbook of Epistemology*. Dordrecht: Kluwer Academic Publisher.; Longino, Helen E. 1993. Feminist standpoint theory and the problems of knowledge. *Signs: Journal of Women in Culture and Society*, **19**(1): 201-212.; Pohlhaus, Gaile. 2002. Knowing communities: an investigation of Harding's standpoint epistemology. *Social Epistemology: A Journal of Knowledge*, **16**(3): 283-293.; Harding, Sandra. 1995. Strong objectivity: A response to the new objectivity question. *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science*, 104(3): 331-349.

have had on Aboriginal self-determination.[184] This commentary demonstrates that our laws and policies on privacy do not easily accommodate the social importance of privacy that is apparent in many Aboriginal cultures. This lack of attention to collective notions of privacy contributes to the de-valuing of certain groups of people. Instead of concluding that certain groups of people require distinct privacy considerations (as Allen claims that women do) this analysis provides evidence that our common sentiments about privacy and information technologies are deeper than "mere preferences and desires".[185] It is the lack of attention to the social value of privacy that prevents our privacy from being properly secured in this information age and the evidence for this is that our privacy conceptions do not protect the least vulnerable. This lack of recognition of the collective value of privacy threatens everyone's privacy and contributes to the de-valuing of certain groups of people.

Because I am going to expand the following marginalized analysis and apply it to my critique on data mining, it is worth explaining the philosophical relevance of the marginalized viewpoint and how this viewpoint can be a valuable place from which to critique the underlying assumptions of society and fight for social change.

A major contributor to theories on the privileged position of the marginalized is Sandra Harding's feminist standpoint theory.[186] The foundation of this theory is the

---

[184] French, Martin. 2003. Privacy across cultures: Indigenous self-determination and the politics of information. *Unpublished Thesis*. University of Victoria, Victoria, B.C.

[185] Nissenbaum, Helen. 1998. Protecting privacy in an information age: The problem of privacy in public: pp. 580-581.

[186] Pohlhaus, Gaile. 2002. Knowing communities: an investigation of Harding's standpoint epistemology. *Social Epistemology: A Journal of Knowledge*, **16**(3): 283-293., and Harding, Sandra. 1995. Strong objectivity: A response to the new objectivity question. *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science*, 104(3): 331-349.

belief that even though there can never be full and completely objective knowledge of the world, there can be degrees of knowledge that have more or less epistemic truth. With roots in the Marxist tradition, feminist standpoint theory holds that we are informed by our positions and relations in the world. Notably, we define ourselves by the work we do and in turn, the work shapes us. Thus knowledge is relational and shifting as our society changes.

Those who exist in one sphere of society have knowledge within that sphere. Thus those in the dominant class have knowledge of the world as it appears to them through their interactions, projects, relationships, and work. Arguably most of the dominant class sees the world in a certain way, with open doors and easy access. The subjugated, however are often kept on the margins of society, not fully realizing an equal social standing. Furthermore, the marginalized often work for and exist in social relationships that perpetuate their own repression. Those who work and struggle in the lower classes may come to gain insight into how their own repression helps to maintain the current social order; they may come to see that those in the dominant classes are ignorant of or choose to ignore the suffering of those who bear the burdens of the social order. It is this dual relationship that may allow those marginalized to have a privileged epistemic position. They may come to have more insight and a more objective knowledge of the social order than those in the dominant class who see only within their own social position. Armed with a more complete awareness of the social order, those with this awareness can then choose to take a standpoint and educate others and fight for change.

As I have presented it, this theory may appear a bit heavy-handed. One could argue that a person or group can never know what it is *really* like to exist in another social position and thus it is not truly a greater epistemic position that is gained by the subjugated; rather it is that the subjugated gain a greater awareness of how the current social order disadvantages them. In other words, the repressed people in society are in a better position to learn about the negative effects of unfair social policies, since repressed people are the ones who are often disadvantaged by these policies. This less forceful version of standpoint theory is the one that I am employing throughout my thesis. My claim is that in a free and egalitarian society, we aim for fairness and equity and we desire to promote the social good. We want to learn how the current social order disadvantages members of society and how we can remedy such harms. Those that are marginalized can come to have a greater awareness of the unjust workings of society because they are often the ones who suffer the negative consequences. As such, certain marginalized people may be in a better position from which to critique the integrity of social policies and the current social order because they are able to highlight injustices that do not affect the privileged members of society. This is the "better epistemic position" of the marginalized that applies to my ethical analysis of data mining. It is this "softer" version of standpoint theory that pertains to the marginalized analysis I will be presenting.

Before discussing the marginalized account, I need to mention two major points of critique that are often raised in relation to standpoint theory. The first is the apparent assumption of the uniformity of the marginalized in their own position, experiences, and desires; and the second is the apparent confusion over whether or not one must actually

occupy a certain social position in order to "see" critically from that position. In a way, these points can be addressed together.

Harding takes pains to acknowledge that women and marginalized people are not uniform groups and this must be kept in mind. We cannot assume that one marginalized person in society will feel the effects of a social policy the same way as another, or have the same desire for changing the system as another. It is the *standpoint* that one takes that is central and not just the social position to which one belongs; it is how one chooses to think about and act on injustice that is important. Thus, those who are not in a marginalized position can still learn about unjust policies that affect other members of society and can decide to adopt a standpoint against such policies. For example, Harding states that even though a man cannot see through the eyes of a woman, a man can still come to understand a woman's point of view. This being said, it is important to keep listening to those in certain positions and not co-opt their voices. We should acknowledge that those who are disadvantaged by policies may be in a better position to explain the nature of the injustice and how the harms can be mitigated. Thus, when we question unjust social policies, it is very important to listen to those people who have been affected by the injustice, as well as those who have not been disadvantaged but who have come to understand the nature of the inequality.

With these points of caution in mind, I present a critique from the margins to provide evidence that privacy has a collective value that is often overlooked.

Privacy Viewed from the Margins

Martin French examined the impact that dominant privacy assumptions have had on Aboriginal self-determination.[187] Space constraints prevent a full account of French's analysis so I present here a summary of his relevant findings.

Aboriginal peoples in Canada have struggled against cultural genocide, discrimination and subjugation. They have not only been marginalized in social relationships of repression and domination, they have also been physically marginalized through the dissemination of Traditional Lands and the creation of Reservations.

Aboriginal People stand somewhat apart from other minority groups in that they do not want allowances that come from within a Governmental system – they want acknowledgement of their rights that came before this system. French focuses on Health Information and the resulting privacy violations for Aboriginal peoples. He argues that the current conception of privacy is liberal and atomistic.[188] The fair information principles evident in legislation and codes of conduct that have evolved from this conception are based on the notion that everyone in society ought to enjoy relatively the same level of privacy and that privacy is conceived of individually.[189] There is an assumption that everyone, more or less, is under the same amount of informational surveillance and has recourse to relatively the same amount of privacy. As such, legislation has not addressed groups of people who may be monitored at a higher rate or whose information may cause more problems for them.

---

[187] French, Martin. 2003. Privacy across cultures: Indigenous self-determination and the politics of information.

[188] *Ibid.*, p. 73.

[189] *Ibid.*, p. 73.

French also points out that an individualistic or atomistic conception of privacy may actually enable further surveillance and privacy violations.[190] He argues that as people are assigned to groups and differentiated not by themselves, but by a system, their identities become determined by this system and not by themselves. French writes, "If data subjects are conceived of, by information privacy regimes in atomistic terms, their entire social context in which their privacy interests are articulated are erased."[191]

French notes that when information privacy is conceived of in this homogenous way it ignores the fact that information may have different meanings in different settings.[192] As we have seen, your grocery purchases may not mean that much by themselves but coupled with your address and monthly income they may take on a whole new meaning.

It is clear from French's analysis that the lack of attention to the importance of context in privacy conceptions is detrimental to many in society. When privacy conceptions and laws focus solely on the importance that privacy provides to the individual, the social importance that privacy plays is largely ignored. French emphasizes the way that traditional privacy conceptions have overlooked the contextual relevance of the information and the different consequences that certain information might have for certain groups of people. His claim is that a lack of understanding of the importance of context and consequences of information further muddies ethical debates and may work

---

[190] *Ibid.*, p. 74.

[191] *Ibid.*, p. 74.

[192] *Ibid.*, p. 75.

to perpetuate subjugation. It may also increase the level of surveillance for vulnerable individuals.

I am not attempting to draw exactly the same conclusions in the ethical analysis of data mining that I am presenting; I am not claiming that data mining causes certain people to be monitored at a higher rate and my commentary is not focusing explicitly on how the context of information shifts during data mining.[193] Rather, French's analysis is important to my thesis because of the way he illustrates that there is a liberal and atomistic conception of privacy that is functioning in our laws and policies. This conception prevents the recognition of the collective nature of privacy that is explicitly acknowledged in many Aboriginal cultures. Thus, French provides evidence that privacy has a social value that is being ignored and this has lead to the further repression and marginalization of certain groups of people.

Instead of concluding from French's commentary that certain groups of people require distinct privacy considerations (as Allen claims that women do), I take it that his analysis demonstrates that our common sentiments and intuitions about privacy and information technologies are deeper than "mere preferences and desires".[194] I believe that it is a lack of attention to the social value of privacy that prevents everyone's privacy from being properly secured in this information age and the evidence for this is that our privacy conceptions do not protect the least vulnerable. In other words, ignoring the collective value of privacy threatens privacy on the social level and we see the effects of

---

[193] Helen Nissenbaum focus on what she terms the "contextual integrity" of information and the way that such an integrity can be respected or violated. See: Nissenbaum, Helen. 1998. Protecting privacy in an information age: The problem of privacy in public.

[194] Nissenbaum, Helen. 1998. Protecting privacy in an information age: The problem of privacy in public: pp. 580-581.

this when information is collected and used in a way that disadvantages certain groups of people.

French argues that a collective notion of privacy would acknowledge the relevance of the social and shared interests in information. A collective notion of privacy would provide a forum for articulating collective privacy concerns and allow groups of people, especially the marginalized, to have a voice. Building on this social commentary, in Chapter 5 I will develop this notion of collective privacy and demonstrate how it can provide philosophical clarity to the issues raised by data mining. Not only does collective privacy aptly address many of the ethical concerns, it also accounts for the intuitive connections between privacy and data mining applications. I will also examine the role that businesses using personal information play in shaping our collective view of privacy and suggest some policy changes that address the social duties these businesses may have.

Chapter 5:

The Role of Collective Privacy

In the last chapter, I explained that Gavison's conception of privacy as limited access to the person does not provide a framework that easily addresses the ethical problems associated with data mining. Instead of concluding that this technology does not threaten privacy, I explored ways in which our common sentiments about privacy and information technologies are not fully accounted for by this traditional conception of privacy. My claim is that a lack of attention to the social value of privacy has prevented privacy from being properly protected in this information age. As we saw from Martin French's examination, the evidence for this claim lies in the fact that our current privacy conceptions do not protect the most vulnerable. This lack of recognition of the collective value of privacy threatens privacy on the social level and contributes to the de-valuing of certain groups of people.

In this chapter, I will further clarify the concept of privacy as a social value. I develop an analogy of privacy as a public good to address some of the conceptual problems relating privacy to data mining and the associated negative consequences. I then work through the relationship between privacy and data mining to show that acknowledgement of the social value is an important addition to traditional ways of conceiving privacy. To help with this examination, I rely on possible policy solutions that focus on the collective value of privacy. These policies could be powerful tools for social change as they situate our common feelings about the connection between privacy and information technologies while protecting the most vulnerable.

The Analogy of Privacy as a Common Good

From the last chapter, we saw that there are reasons and evidence to support the claim that privacy has an important social value that has been largely ignored in our laws and policies. French's examination of the impact that dominant privacy assumptions have had on Aboriginal self-determination showed us that traditional privacy conceptions have focused narrowly on the importance that privacy provides to the individual. This has prevented a recognition of the larger social value of privacy that is apparent in common sentiments about privacy and information technologies. The effect has been that privacy is not fully protected in our networked society, which has resulted in the de-valuing of certain groups of people. Instead of concluding that certain groups of people require distinct privacy considerations (as Allen claims that women do), I take it that French's analysis demonstrates that our common sentiments about privacy and information technologies are deeper than "mere preferences and desires".[195] It is the lack of attention to the social value of privacy that prevents everyone's privacy from being properly secured in this information age and the evidence for this is that our privacy conceptions do not protect the least vulnerable. This lack of recognition of the collective value of privacy threatens privacy on the social level and contributes to the de-valuing of certain groups of people.

The relationship between privacy, data mining and the associated negative social implications still needs clarification. A useful way to understand these connections comes from Priscilla Regan and her notion that the collective value of privacy is similar to the economists' concept of collective or public goods –

---

[195] Nissenbaum, Helen. 1998. Protecting privacy in an information age: The problem of privacy in public: pp. 580-581.

> "which are those goods defined as indivisible or nonexcludable; no one member of society can enjoy the benefit of a collective good without others also benefiting…If a good is a collective good, then it will not be produced through the market or a market solution will result in suboptimal supply of a collective good. The market is an inefficient mechanism for supplying collective goods."[196]

Though I would not want to say that privacy is the same as a collective good in a strictly economic sense, this analogy is useful for clarifying some of the conceptual difficulties we have encountered thus far. A collective good is one that everyone enjoys in a collective sense.[197] An example is clean air. Everyone shares in and benefits from clean air, thus it is in everyone's interests to promote and maintain clean air. However benefits often accrue to those who exploit collective goods. While some businesses may "do their part" to reduce emissions and promote clean air for all, other businesses may shamelessly pollute the air if it increases their profits. Individuals may also act out of selfish interest and contribute excessively to pollution. This exploitation of the common good benefits the polluters but the costs are distributed to all those who rely on the resource – in this case, everyone. Even though this is an unfair distribution of the benefits and burdens of a collective good, it does not make economic sense for one business to decrease emissions while other factories continue to pollute and make higher profits. At the individual level, someone could argue that she does not need to reduce emissions if no one else does – her contribution to pollution is insignificant in the overall scheme. A tragedy of the commons[198] results as those motivated by their own self-interest decrease the air quality and cause everyone to suffer. Thus the exploitation of

---

[196] Regan, Priscilla M. 1995. *Legislating Privacy: Technology, Social Values, and Public Policy*: p. 227.

[197] *See for example*: Samuelson, Paul A. 1954. The pure theory of public expenditure. *Review of Economics and Statistics,* **36**(4): 387-389.

[198] Hardin, Garrett. 1968. The tragedy of the commons. *Science*, **162**(3859): 1243-1248.

collective goods is detrimental to the common good, and collective goods are therefore often protected by the Government through social policies that structure incentives and penalties designed to secure collective goods such as clean air.

Regan states that currently privacy is not thought of as a collective good but as a private good. It is commonly claimed that people can create their own level of privacy by controlling their own information within the market. The notion is that one can release personal information or withhold it as they choose. For example one can choose to un-list her telephone number, one can choose to become part of a list serve, one can sign-up for a grocery store client card, one can refuse to provide personal information to on-line companies, etc. The idea is that privacy can be divided into components and everyone can establish their own individual level of privacy that works best for them.[199] This claim, that the level of privacy is a personal choice, is commonly advanced by data mining advocates. The argument is often made that if you feel that your level of privacy is threatened by the collection, storage, and use of large amounts of your data then you merely need to prevent your information from being acquired. After all, you are almost always free to opt out of mailing lists and to end business with organizations that collect your information. However Regan notes that viewing privacy as a private good ignores the complexity and interrelatedness of computer and communication infrastructure that makes it extremely difficult to divide privacy.[200] In this way, "Privacy is becoming less an attribute of individuals and records and more an attribute of social relationships and

---

[199] Regan, Priscilla M. 1995. *Legislating Privacy: Technology, Social Values, and Public Policy*: p. 228.

[200] *Ibid.*, p. 230.

information systems or communication systems".[201] Because of this, "If one individual or a group of individuals waives privacy rights, the level of privacy for all individuals decreases because the value of privacy decreases."[202] Likewise, if one person chooses to pollute the air, everyone suffers the consequences. It is not as though one person can waive his rights to clean air and breath pollution, while another person retains her right to clean air and breathes clean air instead; when the collective good of clean air is exploited, everyone suffers the consequences. As I quoted earlier, discussing privacy in a similar context, Bennett and Raab write,

> "it is the *collective*, indivisible value of privacy which is particularly important, and which is perhaps especially vulnerable to attack by inequalities. As a collective value, privacy cannot easily be provided to one person without its being enjoyed by others. Rather like street lighting, to provide for one is to provide for all."[203]

Viewing privacy as a private good ignores the social value that privacy has and its collective nature; privacy is not just valuable to individuals as individuals but it is also an important social value.

We can use the analogy of privacy as a collective good and the example of clean air as a tool to help understand how privacy, data mining, and the associated social costs relate to each other. Privacy as a social good means that everyone shares in the collective level of privacy that is fostered and protected within our society. As we saw, there are common sentiments about the relationship between personal information and privacy. Many feel that the collection, use, and sale of even largely public information can

---

[201] *Ibid.*, p. 229.

[202] *Ibid.*, p. 233.

[203] Bennett and Raab, *supra* note 174 at p. 75. (Bennett, Colin J. and Charles D. Raab. 2003. *The Governance of Privacy: Policy Instruments in Global Perspectives*: p. 41.)

threaten privacy. However personal information has become a profitable commodity and data mining and other information technologies have become ubiquitous. Businesses can exploit privacy as a collective good and profit from the use of personal information while other members of society suffer the costs. Just as people with asthma and respiratory diseases will be the first to suffer the effects of air pollution, the disadvantaged and the marginalized have been the first to suffer the exploitation of privacy by information technologies.[204] This does not mean that the marginalized have unique status and claims with respect to privacy but rather their position is sensitive to the way in which the social value of privacy is not fully accounted for in our current laws and policies. Thus the effects of the exploitation are felt at the individual level, first by the marginalized; but it is the overall collective level of privacy that is threatened when the social value of privacy is not respected.

Returning to the conception of privacy as limited access to the person will help spell out how privacy as a collective good acknowledges the social value of privacy. The previous examination of this definition of privacy through the components of secrecy and anonymity [205] revealed that, traditionally, losses of privacy occur when excessive information is known about a person and/or excessive attention is paid to a person. The way that this definition of privacy focuses on the individual and the connection between the individual and his own information, made it hard to understand how privacy is threatened when the personal information collected from one individual is used to disadvantage another individual. When we think of privacy as a collective good however,

---

[204] How we account for shared interests and how we determine a fair distribution of benefits and burdens is a complicated issue and I will return to discuss it later in this chapter.

[205] The third component in Gavison's privacy definition, solitude, was not relevant to the ethical analysis of data mining because the technology does not threaten one's physical space.

we can understand the connection between one person's information and the negative consequences for another person. It is the social value of privacy that is taken advantage of when personal information is used in this way, not necessarily the privacy of any one individual. It is similar to a situation where the pollution from one city affects the people in another city – the cost of the pollution was displaced. Likewise, the cost of exploiting the social value of privacy is displaced when information about one person is used to create business policies that disadvantage someone else. Thus, it may not appear that data mining and the resulting business policies violate any individual's privacy in the traditional sense; yet we can still understand how the collective level of privacy has been exploited.

Understanding the social value of privacy also helps us to grasp the connection between privacy, individual information and the creation of a detrimental profile. This was conceptually difficult since profiles are usually based on the analysis of vast amounts of de-identified data. It appeared that the use of any individual's information was largely insignificant, and since the information was de-identified, privacy was not a factor. The analogy of privacy as a collective good also helps to clarify these issues when we ground each piece of data in the aggregated effects of collecting and using personal information and the social value of privacy. When a business claims that their contribution to air pollution is insignificant and thus they can pollute all they want, they are choosing to ignore the aggregated effects of air pollution and in this way they are exploiting clean air as a common good. Likewise, the claim that any one piece of personal information is insignificant in making a profile, ignores the aggregated effects that personal information has when consumer profiles are used to categorize and discriminate, and in this way

businesses are exploiting the social value of privacy. Just as each business contributes to pollution, each person's data contributes to profiles. Just as everyone has a stake in shared air, everyone has a stake in the social value of privacy.

Understanding privacy as a collective good also allows us to understand how the larger social costs of data mining are related to privacy and how these costs are properly shared costs and not just individual ones. Privacy conceived as limited access to the person posits the harm of privacy violations on the individual. Even composite or aggregated harms – as in the case where all of the medical records in a hospital were released to the public – would be understood as the harm that it causes to each individual. Again, if we think of clean air, we can think of situations where a group of people would bring about a class action suit because of harm they suffered due to an instance of air pollution – for example, if they all worked at a factory that did not maintain proper ventilation. In this case, the problems would be aggregative or collective – each individual had suffered. We would want compensation for those who became sick and we would want to prevent the problem from happening again lest others get sick. In this way, the aggregate harms felt by these people would also become a societal concern – we want to protect all people from situations such as this. On the other hand, we can think of instances where pollution has decreased the air quality of our environment and this has led to certain vulnerable people becoming sick. In this case, the concern is not just that these people became sick; rather the concern is that the air quality has decreased for everyone. The problem is not properly conceptualized as the aggregated harms suffered by these people; it is a social problem and the telling effect has been that these people

became sick. In this case, all of society has been disadvantaged because clean air as a collective good has been exploited – the quality of clear air has decreased for everyone.

Similarly, the individuals who are disadvantaged by profiling and discriminatory business practices will be the first to suffer the negative social costs associated with data mining but these costs are properly conceptualized as social ones and not individual ones. Denying an individual a cell phone contract because she lives in a certain location disadvantages her, but this is not just an individual concern. Nor is it properly thought of as just an aggregate worry when groups of people, such as the marginalized, are similarly disadvantaged. The concern is rightly a social one because it is the collective value of privacy that has been exploited. All of society has been disadvantaged because the collective level of privacy has been exploited. The people who suffer because of these business policies feel the effects of the larger social problem – privacy has been weakened for all.

This analogy of privacy as a collective good also highlights the fact that the traditional ways of conceiving privacy as valuable to the individual are still very important, and yet why the social value of privacy is not reducible to individual privacy. As we have seen, privacy as an individual value is significant for the benefits it provides to the individual and the societal benefits that flow from this. It is true, as Gavison rightly argues, that privacy conceived of as limited access to the person, is necessary for the functioning of a free and egalitarian society. I am not claiming that we should reject this notion of privacy. Our current laws and policies properly allow people to fight for privacy as a civil liberty and this individual and shared benefit is still very important. Rather, my analysis of the relationship between privacy and data mining has shown that

this traditional conception is no longer comprehensive enough to fully protect privacy in this information age. As we have seen, privacy as limited access to the person focuses on the benefit that privacy provides to the individual. The social benefits of privacy are viewed too narrowly as a by-product of this relationship. Regan discusses this point with reference to Gavison's conception of privacy,

> "Without privacy, it would be more difficult, as Ruth Gavison points out, to develop diversity and social pluralism. But this makes the importance of privacy dependent upon, or a result of, the importance of tolerance or social pluralism. The possibility that individuals have a common interest in privacy beyond its importance to social diversity needs to be explored."[206]

This focus on the individual misses the larger social significance that privacy holds in our networked society. Bennett and Raab write, "the individualist assumptions behind the privacy paradigm have tended to overshadow the question of the distribution of privacy protection (and conversely surveillance) in modern societies."[207] As Allen amended Gavison's conception of privacy, technologies such as data mining are highlighting the problems of securing privacy in our information age; and this may mean that our traditional conceptions of privacy need to be amended to fully protect privacy as a social value.

My analysis has shown that there is a need to consider the larger social role that privacy plays and the impact that technology has had on our collective privacy. As Regan writes,

> "Aligning privacy with societal interests would remove some of the difficult philosophical and policy issues involved in reconciling the balance between individual and society. Recognition that privacy has

---

[206] Regan, Priscilla M. 1995. *Legislating Privacy: Technology, Social Values, and Public Policy*: p. 223.

[207] Bennett, Colin J. and Charles D. Raab. 2003. *The Governance of Privacy: Policy Instruments in Global Perspectives*: p. 7.

> some features of a public or collective good would make clearer the institutional or organizational interests in personal information and the weaknesses of a market solution in providing privacy protection."[208]

In other words, recognizing privacy as a social good provides a more comprehensive understanding of privacy in our networked society. The question is: how do we accomplish this recognition? Do businesses have duties? What about government polices? In this next section, I will explore different ways that the social value of privacy might be acknowledged in our society.

Acknowledging the Social Value of Privacy

Firstly, I must note that agreeing that the social value of privacy should be respected does not mean that the collection and use of personal data must end. Nor does it rule out the trading and selling of information. The collection and use of personal information and data mining analyses do not need to lead to negative social consequences. There are many advantages for businesses and consumers that should not be overlooked. The following policy suggestions are not designed to prohibit the use of information technologies, rather they focus on accountability and how these technologies could be used ethically.

One suggestion is to let businesses amend their own policies and police themselves with respect to the use of personal information. Businesses could acknowledge the social value of privacy by phasing out discriminatory business practices and creating policies that respected the social good. This would avoid Government participation in this issue, which may be appealing to many people who feel that the Government should have minimal involvement in the market economy. However as van

---

[208] Regan, Priscilla M. 1995. *Legislating Privacy: Technology, Social Values, and Public Policy*: p. 231.

Wel and Royakkers have discovered from their research, for the most part businesses that rely on data mining do not think that the technology is ethically problematic.[209] As I have stated, the argument is often made that data mining technology does not threaten privacy and thus the technology is ethically sound. It is possible that elucidation of the connection made here between the collective nature of privacy and the negative social consequences associated with data mining may persuade organizations to reconsider their policies, however that seems unlikely. As van Wel and Royakkers show, many businesses claim that if privacy is a concern for customers, laws and policies already exist to protect personal information.[210] The authors conclude that self policing is not a very promising solution.

In addition, self policing with respect to the social value of privacy could easily lead to a "tragedy of the commons"[211] type scenario. Since it can be profitable for businesses to create discriminatory polices there would not be much incentive for this practice to end. If some businesses continued to categorize customers and profit there would be even less incentive for other businesses to amend their own polices. Thus, it is hard to see how the exploitation of the social value of privacy would cease if businesses were left to police themselves.

Van Wel and Royakkers suggest instead that individuals should work hard to protect their own information and they offer ways that technology might advance to give consumers more control.[212] However, this suggestion places individuals in an overly

---

[209] van Wel, Lita, and Lambèr Royyakkers. 2004. Ethical issues in web data mining.

[210] *Ibid.*, p. 134.

[211] Hardin, Garrett. 1968. The tragedy of the commons. *Science*, **162**: 1243-1248.

[212] van Wel, Lita, and Lambèr Royyakkers. 2004. Ethical issues in web data mining.

burdensome position. Consumers would need to become extremely informed and vigilant in order to keep track of the ubiquitous and subtle uses of information technologies, and the way that data mining can impact society. One could argue that it is unrealistic to assume that individuals could even detect all of the ways in which personal information is collected and used within the business sector. More importantly, this suggestion does not acknowledge the social value of privacy, as it places the burden on each individual to protect their own information. If individuals decide to "give up" privacy, either because they are not fully informed or they do not wish to be constantly on guard, then others may suffer the consequences. Business could still create detrimental policies based on the information they collect. In this way, no one is really protected by such a policy and the most vulnerable will be further disadvantaged. Placing the burdens on individuals also does not acknowledge the benefits that businesses gain from using personal information and the social responsibilities these benefits may entail. Similarly, Regan writes, "By placing the burden on the individual, there is less need to evaluate whether organizational interests are indeed social interests or whether individual privacy interests could be conceived as social interests."[213] The possible burdens that businesses may have in recognition of the social value of privacy is an issue I will address more fully later. For now, it is clear that placing the burden on individuals would not be a solution that acknowledges the collective value of privacy.

It appears that the responsibility to acknowledge the social value of privacy cannot be placed solely on businesses or on individuals. Just as clean air is best securable by social means, it looks as though the collective value of privacy is also best securable by

---

[213] Regan, Priscilla M. 1995. *Legislating Privacy: Technology, Social Values, and Public Policy*: p. 219.

social means. Of course how the concerns that I have presented could be properly addressed in a social policy is a complicated issue.

A tool to help in these considerations is suggested by Regan. She "redefines" privacy so as to articulate how it can be acknowledged as a collective good. As we shall see, I do not take Regan to be defining privacy here in the conceptual sense; I interpret her definition to be a means to show how privacy would function in our laws and policies. Thus I take it that she is not dismissing the value and use of our current conceptions of privacy; her suggestion is a normative one, designed to acknowledge the social value of privacy. She writes:

> "Instead of defining privacy as the right of the individual to control information about and access to himself or herself, privacy would be defined as the right of a society to require institutions using personal information to do so in a manner that respects the shared interests in that information. Policy discussion would then focus both on how institutions are using information and on the common interests and concerns individuals have in that information."[214]

This short policy suggestion is powerful for a number of reasons. Firstly, it does not place the burden of protecting personal information on individuals. The individual is not saddled with linking her information to a larger negative social outcome and she does not need to be responsible for constantly monitoring how information is collected and used in our market economy. Instead, the businesses that benefit from the use of personal information would have to show how it is that this use reflects the shared social interests in that information.[215]

In relation to this point, Regan's policy suggestion properly explains that privacy is not just a right of the individual but also "a right of society". Again, I think that

---

[214] *Ibid.*, p. 232.

[215] The term "shared interests" is one that needs clarification and I discuss it in detail later in the chapter.

privacy is an important individual right; however, articulating the policy this way allows us to understand the collective importance of privacy, especially in this information age. Using the phrase that privacy is "a right of society" also helps make sense of the feeling that a common or collective privacy violation has occurred when the shared interests in information are not respected; in other words it helps us make sense of our sentiments that "our privacy" has been lessened or eroded by information technologies; such as when detrimental business policies disadvantage groups of people and weaken our collective state of privacy.

This being said, Regan's policy would not necessarily work to prevent data collection and use and/or even the sale of data. Rather it would require businesses to be accountable for such collection and use. She places the emphasis on the right of society to require businesses to be respectful and, in this way, her policy does not place the responsibility on businesses or organizations to be the ones to safeguard the social value of privacy; as we have seen, self-policing on the part of businesses is not a very promising proposal. Rather, Regan's policy recognizes the role that businesses play in shaping the collective value of privacy and recognizes that it makes sense that they should be accountable for their influential social role.

Acknowledging that businesses and organizations may have a duty to account for their use of personal information also acknowledges the role that information technology now plays in the marketplace and the economy. As we have seen, information technology is subtle and ubiquitous and even though many data mining advocates state that you are free to withhold information or opt out of policies, etc. it would greatly change the face of our economy if customers did opt out en masse. Acknowledging the

duty that businesses have rightly places the burden of respecting the social value of privacy at the feet of those who benefit from the use of personal information. It makes sense that those who benefit should shoulder the burdens.

Most importantly, Regan's articulation of the claim that there are shared social interests in privacy that need to be respected appropriately addresses the major ethical issues that I have presented in this thesis. If a business is using consumer profiling to fuel questionable business practices such as customer firing and weblining, then they are not respecting the shared interests in the information that they collected or bought. After all, questionable business practices have never really respected larger social interests and the fact that data mining can now be used to make these practices easier should be addressed. Businesses would also be responsible for detrimental policies that were based on profiles built from inaccurate and incomplete information, as most profiles are these days. They would have to account for any misleading profiles that used information in a way that does not respect the shared interests in that information.

Regan's policy also addresses business practices that are not so outwardly questionable. As I have explained, even if it were possible to create profiles that were based on completely accurate statistics, there are still important ethical considerations. Policies can have far reaching negative social consequences if they contribute to the de-valuing of groups of people. Any business that adopts a policy based on mined information would need to show that the shared interests in the personal information are accounted for even if the policy looks justifiable from a market perspective. Likewise, businesses could not blindly adopt the policies created by other businesses – they

themselves would need to show why adopting such a policy respects the shared interests in the information that was used in the first place.

Regan's notion of collective privacy and shared interests in personal information also addresses the ethical concerns raised by Martin French's critique from the margins. French explained that currently our laws and policies focus on the benefits and protections that privacy provides to the individual. This lack of recognition of the social value of privacy has contributed to the disadvantaging of certain people in society. The idea that the individual can control her own level of privacy as it suits her ignores that privacy has a collective nature; if one individual or group waives their privacy rights then the level of privacy decreases for everyone.[216] Likewise, if one person chooses to pollute the air, everyone suffers the consequences. It is not as though one person can waive their rights to clean air and breath pollution, while another person retains their right to clean air and breathes clean air instead; when the collective good of clean air is exploited, everyone suffers. Discussing privacy in a similar context, Bennett and Raab write, "it is the *collective*, indivisible value of privacy which is particularly important, and which is perhaps especially vulnerable to attack by inequalities."[217] As we saw, data mining and other information technologies rely on personal information that is not protected by privacy policies and laws because it appears that people have consented to the collection and use of their personal information. Yet the use of this information can lead to categorization and coding that may perpetuate stereotypes. If business profiles become social categories they may be interpreted as more telling than individual data. If it

---

[216] Regan, Priscilla M. 1995. *Legislating Privacy: Technology, Social Values, and Public Polic:* p. 233.

[217] Bennett, Colin J. and Charles D. Raab. 2003. *The Governance of Privacy: Policy Instruments in Global Perspectives*: p. 41.

appears that prejudices are endorsed, the people affected may be prevented from attaining full autonomy since they will be judged by identities imposed upon them by others, instead of by themselves. These disadvantaged people are feeling effects of the exploitation of privacy as a social good. Failing to acknowledge the collective value of privacy threatens privacy for everyone and contributes to the de-valuing of certain groups of people. Regan's notion of collective privacy recognizes that there are shared interests in personal information that need to be respected. These shared interests would be respected by business policies that do acknowledge the diversity of our society and the relevance that the collection and use of even public information can have to various people.

Thus Regan's policy is promising for many different reasons, yet as positive as this policy is, there are some worries. One issue is that the term "shared interests" needs qualification. It raises epistemological concerns, such as: how would collective privacy interests be determined and measured? There are also ethical concerns: for example, there are reasons to think that the interests promoted by a collective privacy policy would be those of the dominant class and the interests of the marginalized would still be overlooked. Firstly, as we have seen, the notion of "shared interests" is grounded in acknowledging the collective value of privacy and thus this terminology provides a powerful context for the marginalized groups to make changes. The connections between privacy, personal information, and negative social consequences provide a framework for groups of people to voice concerns over polices that ignore their collective interests in personal information, even if that information is not solely their own. Thus, even if policies were not thoughtfully implemented with the concerns of marginalized groups in

mind, once it was known that the interests of these people had not been respected, the groundwork would already be in place to make changes.

Another way to think about the term "shared interests" is in terms of the analogy of privacy as a collective good. Collective goods like clean air are often protected with social policies that have to balance competing interests. In most cases the aim is to distribute the burdens and benefits of collective goods "fairly", not "equally" per se. Thus businesses and individuals may not get their way if they want lenient policies that support their own self-interests while disadvantaging others. Economic factors must be given consideration and the interests of the public, especially the vulnerable members of society, need to be considered. More abstract measures such as future concerns and the betterment of society also play into decisions. This balancing act is not without complications. Each of us belongs to a variety of fluid groups and are impacted by different social policies in different ways at different times. As such, social policies often need to adapt to changes in the social order and adjustments need to be made when new information becomes relevant. This difficult process of protecting collective goods is not without its problems; however the aim is to promote the social good.

Likewise, with respect to the "shared interests" in personal information, differing interests would need to be weighed so that the social value of privacy would be respected. Here the "shared interests" would include the economic interests of the businesses that rely on data mining; the interests of consumers who benefit from data mining applications and the collection of information (i.e.: when discounts and incentives are provided in exchange for personal information); the interests of those who are disadvantaged by data mining applications; and the more abstract social interests such as the negative social

consequences over time and the betterment of the social good. As with clean air, interests would not necessarily count equally. A fair distribution of benefits and burdens of a collective good means that some people must "sacrifice" more than others. Health risks are often given more weight in a decision than economic benefit. Likewise, serious social outcomes and the de-valuing of certain groups of people should be given more weight than economic gain. The economic factor would still be a major consideration, however, as information technologies are part of the fabric of our economy. Thus, the goal of determining the nature of the "shared interests" in personal information would not be to stop data mining and other information technologies. Rather, as with collective goods that are protected in our society, the balancing of interests and determining a fair distribution of the benefits and burdens with respect to the social value of privacy would need to focus on the social good in all its facets.

Even with this complication, Regan's policy helpfully articulates the connections between privacy, data mining technology, and the associated negative consequences in a normative way. As such I think that it is a very valuable contribution to the discourse on this subject and could be used to make some appropriate changes. Regan has laid some important groundwork, yet it is not easily apparent how such a policy would operate in our current society and how the collective value of privacy and the "shared interests" in personal information could be properly respected. It appears that a social policy designed to acknowledge the social value of privacy would be difficult to implement. However, we already have social policies in place that address many of the issues outlined in this

thesis.[218] For example, we have laws that address discrimination and human rights violations and focus on giving a voice to the vulnerable members of our society. Such laws could provide a basis for policies that discourage business policies that de-value and subjugate people. We also have laws that focus on informed consent in the medical setting and these could be used to address the issues of coercion, ambiguous privacy policies, and whether the consumer is "freely" consenting to provide information. Patent, copyright, and property rights could also be blended to address the notion of the right to say what is done with personal information even if it is public and de-identified. In this way, we could construct a policy that is not entirely new but based on current policies in the way society has structured protection of various rights and promotion of the social good.

It may look from this commentary that I am now rejecting the notion of privacy as a central issue in this ethical analysis; perhaps it seems that the issues in this thesis are really reducible to discrimination, informed consent, property rights, etc. This is not the case at all. These policies are a way to frame the concerns of the social value of privacy in a way that protects the collective level of privacy that has been weakened in our networked economy. Each of these policies could contribute to the larger policy and these policies highlight that the considerations raised by this analysis are neither unique nor unheard of. The social value of privacy accords with our common sentiments and the intuitive connection we feel between privacy and information technologies. A notion of the collective value of privacy is not reducible to these issues; rather they show us that we

---

[218] This idea originated from some commentary by Nissenbaum, though I depart from her analysis. Nissenbaum, Helen. 1998. Protecting privacy in an information age: The problem of privacy in public: p. 594.

already have the tools to address these issues. As I stated, it is difficult with our current "privacy vocabulary" to articulate the relationship between data mining and privacy.[219] This is due, in part, to the fact that philosophical privacy discussions have centered on abstract justifications for privacy whereas the privacy concerns of today that are expressed in common sentiments are derived from real circumstances with real consequences.[220] Thus we must look beyond the traditional ways of articulating privacy. The sentiments of the aforementioned policies, when taken together, illuminate the relationship between privacy and the social implications that stem from the collection and use of large amounts of personal information. I hope that this thesis helps lay the foundation for others to explore this connection more deeply.

In conclusion, I would like to re-state that I do not think that data mining is problematic in-and-of itself; within the business community, it has many legitimate and beneficial uses. However data mining and other information technologies cause ethical concerns when they are used to categorize and discriminate. I have shown that even though there is an intuitive connection between privacy and personal information, it is hard to conceptualize the troubling issues raised by certain data mining applications in terms of privacy. This is largely due to the emphasis that traditional privacy definitions place on the value of privacy to the individual. The notion of collective privacy emphasizes the broader social importance of privacy and provides philosophical clarity to the privacy issues raised by data mining. The policy suggestions that result from acknowledging the social value of privacy could benefit many in our society and work to fortify our privacy in this information age.

---

[219] Vedder, Anton H. 1999. KDD: The challenge to individualism: p. 280.

[220] Regan, Priscilla M. 1995. *Legislating Privacy: Technology, Social Values, and Public Policy:* p. 224.

Literature Citied

Allen, Anita. 2003. Privacy. In Hugh LaFollette (ed.), *The Oxford Handbook of Practical Ethics.* Oxford: Oxford University Press.

Allen, Anita L. 1984. Women and their privacy: What's at stake? In Carol C. Gould (ed.), *Beyond Domination.* Totawa: Rowman & Allanheld.

Austin, Lisa. 2003. Privacy and the question of technology. *Law and Philosophy,* **22:** 119-166.

Bennett, Colin J. and Charles D. Raab. 2003. *The Governance of Privacy: Policy Instruments in Global Perspectives.* Hampshire, UK: Ashgate Publishing Limited.

Bennett, Colin J. 2001. Cookies, web bugs, webcams and cue cats: Patterns of surveillance on the world wide web. *Ethics and Information Technology,* **3:** 197–210.

Bloustein, Edward J. 1964. Privacy as an aspect of human dignity: An answer to Dean Prosser. *New York University Law Review*, **39:** 962-1007. In, Ferdinand Schoeman (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology.* (pp. 156-202) New York: Cambridge University Press.

Bok, Sissela. 1983. *Secrets: On the Ethics of Concealment and Revelation.* New York: Pantheon Books.

Case, Donald O. 1998. The ethics of caller identification services. *Journal of Information Ethics.* **7**(1): 24-35.

Cavoukian, Ann. 1998. Data Mining: Staking a Claim on Your Privacy. *Information and Privacy Commissioner's Report.* Ontario, Canada.

Colleen, Angel. 2000. The right to privacy. *Journal of Information Ethics*, **9**(2): 11-25.

Custers, Bart. 2001. Data mining and group profiling on the internet. In Anton Vedder (ed). *Ethics and the Internet.* Antwerpen-Groningen-Oxford: Intersentia.

Danna, Anthony, and Oscar H. Gandy Jr. 2002. All that glitters is not gold: digging beneath the surface of data mining. *Journal of Business Ethics*, **40**(4): 373-386.

DeCew, Judith Wagner. 1997. *In Pursuit of Privacy: Law, Ethics, and the Rise of Technology.* Ithaca: Cornell University Press.

Fayyad, U., G. Piatesky-Shapiro and P. Symth. 1996. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, **39**(11): 27-34.

Frawley, W.J., G. Piatetsky-Shapiro and C.J. Matheus. 1991. Knowledge Discovery in Databases: An Overview. In G. Piatetsky-Shapiro and W.J. Frawley (eds), *Knowledge Discovery in Databases*. Menlo Park, Cal., Cambridge, Mass/London: AAAI Press/MIT Press.

French, Martin. 2003. Privacy across cultures: Indigenous self-determination and the politics of information. *Unpublished Thesis*. University of Victoria, Victoria, B.C.

Fried, Charles. 1968. Privacy [A moral analysis]. *Yale Law Jounal*, **77**: 475-493. In, Ferdinand Schoeman (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*. (pp. 203-222) New York: Cambridge University Press.

Fulda, Joseph S. 2001. Data mining and privacy. In Richard A. Spinello and Herman T. Tavani (eds). *Readings in CyberEthics*, Jones and Bartlett.

Gandy, Oscar H. Jr. 2000. Exploring identity and identification in cyberspace. *Notre Dame Journal of Law, Ethics, and Public Policy*, **14**(2): 1085-1111.

Gavison, Ruth. 1980. Privacy and the limits of the law. *Yale Law Journal*, **89**: 421-471. In, Ferdinand Schoeman (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*. (pp. 346-402) New York: Cambridge University Press.

Goold, Benjamin J. 2002. Privacy rights and public spaces: CCTV and the problem of the unobservable observer. *Criminal Justice Ethics*, **21**(1): 21-27.

Hardin, Garrett. 1968. The tragedy of the commons. *Science*, **162**: 1243-1248.

Harding, Sandra. 1995. Strong objectivity: A response to the new objectivity question. *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science*, 104(3): 331-349.

Harnad, Stevan. 2005. To cognize is to categorize: Cognition is categorization. In Henri Cohen and Claire Lefebvre (eds.), *Handbook of Categorization in Cognitive Science*. Amsterdam: Elsevier.

Huge database of phone calls a hidden trove of behaviors. 2006, May 12. *The Globe and Mail*. p. A11.

Lennon, Kathleen. Feminist epistemology. In Ilkka Niiniluoto (ed.). 2004. *Handbook of Epistemology*. Dordrecht: Kluwer Academic Publisher.

Longino, Helen E. 1993. Feminist standpoint theory and the problems of knowledge. *Signs: Journal of Women in Culture and Society*, **19**(1): 201-212.

Murphy, Robert F. 1964. Social distance and the veil. *American Anthropologist*, **66**: 1257-1274. In, Ferdinand Schoeman (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*. (pp. 34-35) New York: Cambridge University Press.

Nissenbaum, Helen. 1998. Protecting privacy in an information age: The problem of privacy in public. *Law and Philosophy*, **17:** 559-596.

On the Record: Scott McNealy. 2003, September 14. *San Francisco Chronicle.* Retrieved January 14, 2006 from: www.sfgate.com/cgi-bin/article.cgi?file=/chronicle/archive/2003/09/14/BU141353.DTL&type=business

Peterson, Gregory R. 2003. Demarcation and the scientistic fallacy. *Zygon: Journal of Religion and Science*, **38**(4): 751-761.

Pierik, Roland. 2001. Group profiles, equality, and the power of numbers. In Anton Vedder (ed). *Ethics and the Internet*. Oxford, Antwerpen-Groningen: Intersentia: 105-123.

Pohlhaus, Gaile. 2002. Knowing communities: an investigation of Harding's standpoint epistemology. *Social Epistemology: A Journal of Knowledge*, **16**(3): 283-293.

Prosser, William L. 1960. Privacy [a legal analysis]. *California Law Review*, **48:**383-423. In, Ferdinand Schoeman (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*. (102-155) New York: Cambridge University Press.

Rachels, James. 1975. Why privacy is important. *Philosophy and Public Affairs*, **4**:323-333. In, Ferdinand Schoeman (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*. (pp. 290-299) New York: Cambridge University Press.

Regan, Priscilla M. 1995. *Legislating Privacy: Technology, Social Values, and Public Policy*. Chapel Hill: University of North Carolina Press.

Reiman, Jeffrey H. 1976. Privacy, intimacy and personhood. *Philosophy and Public Affairs* **6**: 26-44. In, Ferdinand Schoeman (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*. (pp. 300-316) New York: Cambridge University Press.

Robinson, W.S. 1950. Ecological correlations and the behavior of individuals. *American Sociological Review,* **15:** 351–357.

Rueker, Stan. 2006, January 12. *Crystalizing the Process: the Clear Browser for Data Mining Humanities Computing*, Talk at the University of Victoria, Victoria, B.C.

Samuelson, Paul A. 1954. The pure theory of public expenditure. *Review of Economics and Statistics,* **36**(4): 387-389.

Schoeman, Ferdinand. 1984. Privacy: Philosophical dimensions. *American Philosophical Quarterly*. **21**(3): 199-213.

Schoeman, Ferdinand (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*. New York: Cambridge University Press.

Sipior, J.C., B.T. Ward and S.M. Rainone. 1998. Ethical management of employee email privacy. *Information Systems Management*, **15**(1): 41–47.

Stephen, James Fitzjames. 1873. *Liberty, Equality, and Fraternity*. New York: Henry Hold and Co.

Sweeney, Latanya. 2002. *k*-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems,* **10**(5): 557-570.

Sweeney, Latanya. 2000. *Uniqueness of Simple Demographics in the U.S. Population, LIDAPWP4.* Carnegie Mellon University, Laboratory for International Data Privacy, Pittsburgh, PA. Forthcoming book entitled, *The Identifiability of Data.*

Tavani, HermanT. 1999. KDD, data mining, and the challenge for normative privacy. *Ethics and Information Technology*, **1**: 265-273.

Tavani, HermanT. 1999. Informational privacy, data mining, and the internet. *Ethics and Information Technology*, **1**: 137-145.

Thomson, Judith Jarvis. 1975. The right to privacy. *Philosophy and Public Affairs*, **4**: 295-314. In, Ferdinand Schoeman (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*. (pp. 272-289) New York: Cambridge University Press.

van Wel, Lita, and Lambèr Royyakkers. 2004. Ethical issues in web data mining. *Ethics and Information Technology*, **6:** 129-140.

Vedder, Anton H. 2001. KDD, Privacy, Individuality and Fairness. In Richard A. Spinello and Herman T. Tavani (eds). *Readings in CyberEthics*, Jones and Bartlett.

Vedder, Anton H. 1999. KDD: The challenge to individualism. *Ethics and Information Technology*, **1**: 275-281.

WalMart Privacy Policy. Retrieved March 4, 2006 from http://www.walmart.com/catalog/catalog.gsp?cat=538446

Warren, Samuel D., and Louis D. Brandeis. 1890. The right to privacy [The implicit made explicit]. *Harvard Law Review*, 4: 193-220. In, Ferdinand Schoeman (ed.).

1984. *Philosophical Dimensions of Privacy: An Anthology*. (pp. 75-103) New York: Cambridge University Press.

Westin, Alan. 1967. The origins of modern privacy, from *Privacy and Freedom*. New York: Atheneum press. In, Ferdinand Schoeman (ed.). 1984. *Philosophical Dimensions of Privacy: An Anthology*. (pp. 56-74) New York: Cambridge University Press.