University of Alberta

An Exploratory Analysis of the State of Online Privacy

By

Ian Reay ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of Master of Science

Department of Electrical and Computer Engineering

Edmonton, Alberta

Fall 2007

Your file Votre référence
ISBN: 978-0-494-33331-0
Our file Notre référence
ISBN: 978-0-494-33331-0

Abstract

This thesis is composed of a series of exploratory studies which attempt to provide insight into the current state of Online Privacy protections. The first study analyzes adoption characteristics of the Internets standard privacy protection technology known as the Platform for Privacy Preferences (P3P) protocol which allows companies to post machine readable versions of their human readable privacy policy on their websites. The second study analyzes how cultural differences may affect the adoption of privacy technologies. The third study investigates the effectiveness of privacy legislation. Finally, the fourth study investigates the effectiveness of the P3P protocol from and end users perspective. A unifying element these four studies is the usage of the Platform for Privacy Preferences (P3P) protocol as the primary vehicle of exploration into these diverse topics. Results of the thesis suggest that the adoption and maintenance of the P3P protocol is limited and its overall effectiveness as a privacy protection mechanism is likely poor. Furthermore, adoption of the technology appears to vary significantly between cultures and current privacy legislation appears to have had little affect on the actions or organizations. The studies do present a novel use of the P3P protocol as a means for large scale automated analysis of the stated actions of the Internets websites. To date, such an analysis has been impossible due to the limited structure existing in human readable privacy policies which are the Internets defacto privacy protection mechanism.

Acknowledgements

We wish to thank Alexa.com for graciously providing us with a copy of their Top 100,000 list.

Contents

Tables

Figures

Equations

1 General Introduction

With the growth of pervasive computing technologies, personal information has never been more accessible, persistent, or valuable. A recent report by The Canadian Internet Policy and Public Interest Clinic [1] describes how an industry has developed to share this information which can provide marketing insight and business intelligence to organizations throughout North America. This marketplace is composed of both for profit and cooperative organizations who collect, aggregate, analyze, and disseminate information. The information they manage originate from a variety of sources including magazines, newspapers, mail order retailers, payment processing companies, and online services [1]. This information can include personal credit/purchase histories, payment information, interests, and tendencies often compiled into personal profiles. Direct marketers are usually the recipients of this information and use it to enhance the efficiency of marketing campaigns since accurate personal profiles can provide marketers with the ability to target individuals or groups of consumers. These actions can also benefit consumers since marketing programs may become more efficient resulting in lower purchase prices for consumers. However, it is inevitable that information will be leaked, stolen, or sold to individuals who will use it inappropriately. Examples include identity theft [2], fraud [2], differential pricing [3], and offensive/intrusive marketing campaigns. The Federal Trade Commission has valued the cost of fraud alone at $50 Billion dollars in 2003! [2]. Obviously, studies into privacy and information security are of high value, but they are difficult to undertake due to the interdisciplinary nature of the privacy topic. Researchers must be aware of differing cultural privacy perceptions, how individual privacy perceptions vary, legal privacy protections, data mining techniques, how pervasive computing technologies enable information collection, data aggregation methods, information warehousing methods, how businesses processes benefit from sensitive information, and information security concepts.

In an effort to provide guidance to governments and researchers developing privacy protection solutions, the OECD has developed recommended guidelines for the protection of privacy and international information flows [4]. These guidelines exist as seven privacy principles and include:

1) Collection limitation principle: Collection of information should be limited and with the consent of the information subject.
2) Purpose specification principle: Information should be accurate, complete, and up-to-date.
3) Use Limitation principle: Information should not be disclosed unless the information subject consents or otherwise lawfully allowed.
4) Security Safeguards principle: Reasonable security safeguards should exist to protect against theft, destruction, use, modification, or disclosure of information.
5) Openness Principle: Information subjects should be able to determine what personal information an information collectors possesses.
6) Individual Participation principle: Information subjects should be able to access personal information held by information collectors.
7) Accountability Principle: Information collectors should be held accountable to the above principles.

Principles 1, 2, 3, 5, and 6 attempt to reduce the information asymmetry between the information collectors, and information subjects thereby empowering information subjects with the ability to determine what consequences may arise from the sharing of information. Human readable privacy policies (HRPP's) are the most widely adopted approach for satisfying these principles. These documents are meant to inform website patrons about what information will be collected, how it will be used, who it may be shared with, and how long it will be retained for. Usually these documents are recognized as legally binding agreements between websites and their patrons and as a result, they can be lengthy [5] and convoluted to understand since they often contain obscure legal terminology beyond the reading comprehension of average Internet users.

In view of these significant short comings, the World Wide Web Consortium (W3C) has created the P3P protocol. This protocol aims to solve many of the deficiencies of HRPP's by providing a procedure for encoding some of the contents of HRPP's into a machine readable format. Software agents working on the users behalf request these documents, analyze their contents, and display relevant information to the user. It is hoped that these software agents can reduce the cognitive demands placed upon end users by filtering extraneous information, thereby allowing them to make better decisions.

This thesis uses the P3P protocol as a vehicle for exploration of the complex topic of online privacy. The major questions that will be investigated are:

1) Is the P3P protocol being widely adopted by organizations?
2) How have legal influences affected P3P adoption and organizational actions?
3) How have cultural variations affected P3P adoption and organizational actions?
4) Can P3P be considered an effective mechanism in reducing the information asymmetry between information subjects and information collectors?

This will be accomplished through a series of studies undertaken over a period of two years. The first study (Chapter 2) is composed of an investigation of the adoption and maintenance of the P3P protocol by Internet websites. Adoption of the P3P protocol by websites is extremely important since the P3P protocol relies upon websites to post these policies on their websites.

The second study (Chapter 3) analyzes the information contained within P3P policies to determine how cultural differences in privacy perceptions affect the adoption of the P3P protocol and the contents of posted policies. Current technological proposals for privacy protections fail to take into account how concerns for privacy vary between cultures. In order for privacy protection technologies to be effective, they must accurately target the concerns of its users. To date, most technologies have been developed with a western centric view of the issue. Only recently have arguments have been made that culturally sensitive solutions are required [6].

The third study (Chapter 4) utilizes the information contained within P3P policies to investigate the effectiveness of current privacy protecting legislation. This analysis is important since many privacy initiatives involve the adoption of legislated protections since there is a developing belief that self regulatory and technological attempts to protect privacy have failed.

The final study (Chapter 5) investigates whether the P3P policies posted on websites are ever used by average Internet users. To date, P3P adoption research has discussed only website adoption of this protocol. If P3P is to be considered a realistic method for Internet users to protect themselves, information indicating that users are actively using the technology is required.

Once complete, Chapter 6 will provide a summary, conclusions, and proposals for future work. Due to the complex nature of the problem of privacy, this thesis does not claim to be an exhaustive analysis of all privacy issues on the Internet. Instead, this thesis attempts to improve upon and extend previous studies through the development of automated techniques for analyzing the P3P protocol. Further, the usage of the P3P protocol as a vehicle for analyzing the effectiveness of legal protection is both novel and allows for analysis on a scale far greater that previous studies which were severely limited in their scope due to the lack of automated analysis methods.

1.1 Bibliography

[1] "On the Data Trail: How Detailed Information About You Gets Into The Hands of Organizations With Whom You Have No Relationship," The Canadian Internet Policy and Public Interest Clinic, Ottawa, Ontario, Canada April 2006.

[2] Synovate, "Federal Trade Commission - Identity Theft Survey Report," September 2003.

[3] Andrew Odlyzko, "Privacy, Economics, and Price Discrimination on the Internet", Advances in Information Security, Vol 12, Springer US, 2004.

Adoption and Maintenance of the P3P Protocol

2 Chapter 2 Introduction

The ease of data creation, manipulation, and aggregation in the online world has removed many of the previous barriers that protected a user's privacy [1-3]. Users are now confronted with unfamiliar issues such as having their online actions tracked and profiled [2, 4] and having their personal information sold to third parties without their knowledge [5, 6]. These topics are discussed in a report by the Canadian Internet Policy and Public Interest Clinic which describes the current state of information sharing in North America [7]. The report describes an industry active in selling, aggregating, analyzing, and enhancing customer information. These actions allow companies to enhance business efficiency and undertake more efficient marketing campaigns by allowing them to more accurately target individuals or groups. This information, when used inappropriately, can also have a significant effect on the personal privacy of data subjects. For instance, such information has been used to create psychological profiles [8], terrorist profiling [5], and identity theft [9].

In an effort to provide guidance to companies about what types of data collection is reasonable, privacy principles have been developed by several governments and organizations. Examples of such principles include the OECD's privacy principles [10], the FTC's fair information practice principles [11], the U. S. Safe Harbor principles [12] and the principles contained within the European Unions Data Directive [13]. Since the OECD's principles have been adopted by all 30 member nations including the US and many EU member states, these principles have the widest international support. These principles provide guidance to organizations and their patrons regarding: limiting collection of information, data quality, purpose for collection, limiting information use, security safeguards, and openness regarding the data collector, data-subject participation, and accountability of the data collector. In an effort to assist both organizations and patrons in abiding by and enforcing these principles, a number of technologies and methods have been either proposed or developed [13-22].

This paper analyzes the current state of adoption of one of these technologies; the Platform for Privacy Preferences (P3P) project [14]. P3P has been proposed as a solution to the many problems which exist with human readable privacy policies. These policies attempt to inform website patrons about privacy sensitive actions undertaken by the website. However, half of all Internet users may not have the required education to properly understand the legal terminology used [16, 23] or technologies discussed (such as cookies) [24] within these policies. In an effort to rectify these deficiencies, P3P, an industry-sponsored technology adopted by the W3C as an official recommendation, was developed to provide a mechanism to lessen the cognitive demands made upon Internet users. This is done through the usage of P3P user agents which first locate, then analyze P3P documents posted on a server. Once analyzed, any relevant information can be displayed to the user in an understandable manner.

A symbiotic relationship exists between P3P agents and P3P documents where neither provides value without the other. Lack of P3P agent adoption in web browsers is of paramount concern. The quality of Microsoft's Internet Explorer's default P3P agent has been questioned [25] and Mozilla's Firefox browser does not contain a P3P agent [26]. In addition, a number of critiques have outlined concerns with both the scope of P3P [19, 27-29] and what may be implied when P3P is used in it's most minimal form [30]. P3P may also suffer from an unwillingness or inability of users to make privacy-preserving decisions [31, 32] and semantic inconsistencies [33]. Even through these concerns exist, Linn has stated in early 2005 that he believes P3P is poised for major growth [34] and recently, the W3C released a public working draft of P3P 1.1 in February 2006 [35].

Due to P3P's symbiotic dependence, server-side adoption of the protocol is vital to P3P's viability. Cranor, Byers, and Kormann conducted Web surveys of P3P server-side adoption in May [36] and July 2003 [37]; hereafter, we will refer to the July 2003 survey as the AT&T survey. This survey claimed that P3P 1.0 [38], which became an official W3C recommendation in 2001, was experiencing an increase in adoption by Internet websites. To our knowledge, no further studies have been published since July 2003. Our goal is to provide evidence depicting P3P's current state of adoption, changes in P3P adoption between July 2003 and November 2005, and P3P policy maintenance practices. This will be accomplished through comparisons between the AT&T survey, a pilot survey completed in February 2005, and our full survey completed in November 2005. Our results indicate that most areas of the Internet have seen P3P adoption stagnate. P3P policies frequently contain syntactic errors, and very few websites ever perform maintenance on their policies. In addition, our results indicate that P3P adoption is correlated with website popularity and P3P adoption varies significantly between languages and countries.

The remainder of this chapter is organized as follows. In Section 2.1, we introduce the P3P protocol and review the available P3P user agents. In Section 2.2, we review the survey methodology of the AT&T survey as well as our extensions and improvements. In Section 2.3, we present our results comparing the AT&T survey to our own. In Section

2.4, we present our analysis of our February 2005 pilot study against our November 2005 full study. In Section 2.5, we present our analysis of P3P document errors and maintenance. In Section 2.6, we present our analysis of P3P adoption across language and national boundaries. In Section 2.7, we present our results examining P3P adoption and error rates for the most popular websites on the Internet. Finally, Section 2.8 offers a summary and discussion of future work.

2.1 The P3P Protocol

P3P's goal is to provide a standardized format that will allow websites to express their privacy practices in a format that can be retrieved and analyzed by software agents working on a user's behalf. P3P 1.0 became an official W3C recommendation in April 2002 [14]. The protocol utilizes XML files, whose semantic meaning is governed by a pre-defined XML Schema provided by the W3C [38]. The schema was designed so that creators of P3P documents could translate most of the semantic meaning of HRPP's into an XML document. It is recommended that P3P document authors use a tool to construct their P3P documents to limit violations of XML and P3P document syntax [14]. The semantic correctness of the HRPP too P3P translation is the responsibility of the individual.

Once the P3P document has been created, organizations are responsible for providing access to these documents via their website. The P3P specification uses a policy reference file to direct software agents to the applicable P3P policy. Software agents can locate this policy reference file using one of four possible methods. The first and recommended way is to check to see if the reference file is stored in a well known location on a web server. The path to this well known location is /w3c/p3p.xml. If this location is already reserved for some other use, a link to the reference file can be specified in the HTTP header. The final two methods for resolving a policy reference file is through HTML and XHTML link tags in the provided HTML or XHTML documents.

Once the users agent has resolved the P3P documents location, it will retrieve the document. Once retrieved, it is the agent's responsibility to determine if the P3P document conflicts with the user's previously stated preferences, and display these concerns to the user in a form they understand. It is the responsibility of the software agent's creator to determine what is the best method of displaying information to a user. Some have chosen to using visual and audio queues [39], while others embed textual information into retrieved documents [40]. Figure 2-1 depicts the operation of Internet users, Web servers, and P3P agents when the user requests a document from the server.

The P3P protocol also defines what is known as a P3P compact policy which is not contained in a file but is a character string placed within the HTTP header of server responses. Compact policies were developed to limit the bandwidth needed when

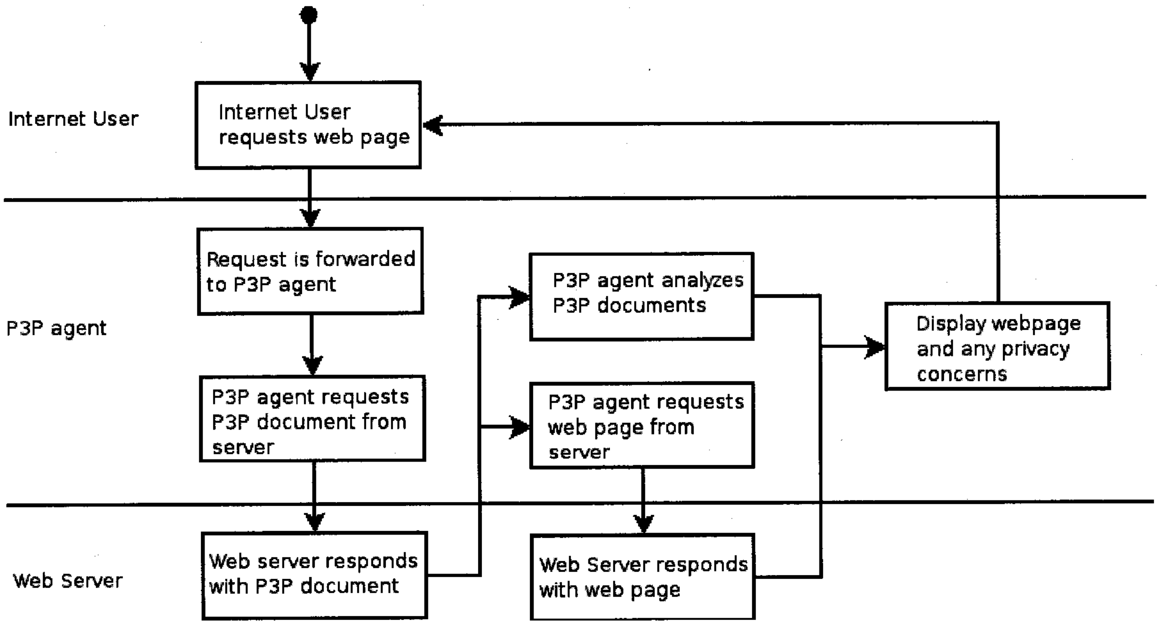
Internet User
Internet User requests web page
P3P agent
Request is forwarded to P3P agent
P3P agent analyzes P3P documents
Display webpage and any privacy concerns
P3P agent requests P3P document from server
P3P agent requests web page from server
Web Server
Web server responds with P3P document
Web Server responds with web page

Figure 2-1: P3P Request Response Sequence

multiple cookies are set, each requiring a different P3P policy, as in the case of third party cookies. Thus, instead of a browser requesting the full P3P policy every time a cookie is set, the browser only checks the HTTP header that contains the cookie in question. The Compact policy will contain only the information relevant to the cookies being set in the transmission. Thus, content of compact policies can be viewed as a subset of the full P3P policy.

2.1.1 P3P User Agents

P3P provides a machine-readable form of a website's privacy policy. A user agent must then analyze the policy content and display the relevant data in a manner that is understandable to the user. However, requirements governing functionality of the P3P agents were not considered in the P3P 1.0 specification. As a result, a wide variety of functionality is provided in default P3P agents and browser extensions.

Microsoft bundled a default P3P agent in Internet Explorer 6.x (IE 6.x). Unfortunately, IE 6.x is limited to blocking cookies and displaying a human readable form of the P3P policy. To our knowledge there is no evidence indicating that IE 6.x's human readable policies are less complex than the usual human readable privacy policies found on websites. Netscape Navigator 7.x's default user agent possessed cookie-handling functionality similar to IE 6.x and presents the human readable policy in a bulleted format [25]. However, this P3P agent is no longer included by default in current Netscape or Mozilla browsers [26].

However, three major P3P browser extensions or proxy services have been created. AT&T's Privacy Bird [39] is an IE extension that was developed by the AT&T survey's authors. A usability study [25] found that this agent generally out-performed IE 6.x's P3P agent and human readable privacy policies. However, they also found the usability of their agent suffered from the lack of available P3P policies. With current P3P adoption rates, users rated the AT&T Privacy Bird agent 2.9 out of 5 where 5 is highly usable (5-point Likert scale). The second browser extension, Privacy Fox [41], was developed as an open-source project for Firefox. The design of Privacy Fox was highly influenced by AT&T Privacy Bird. As of Feb. 2006, the Privacy Fox project appears to be defunct. The third browser extension is the JRC P3P Proxy [40] which filters web requests made through a browser through an intermediate proxy service which requests, analyzes, and returns relevant information to the users browser embedded in the webpage. The most recent release of the JRC P3P Proxy was in June of 2004 and we are currently unable to ascertain the status of this project.

2.2 The Survey Method

The goal of the AT&T surveys was to determine the extent of P3P adoption since its inception as a W3C recommendation in 2001. Our survey followed a similar but improved method in order to determine how P3P adoption has evolved between May 2003 and November 2005. Comparisons will also be undertaken between our February 2005 pilot study and our November 2005 main study. The February 2005/November 2005 comparison allows for a far more detailed analysis due to the limitations in data quality and quantity in the AT&T survey.

In an effort to minimize the threats to validity inherent within the survey, we have adopted the following policies (experimental design practice):

- Implementation bias must be minimized in favor of a faithful implementation of P3P.
- Error analysis must accurately define what constitutes an error in a P3P policy.
- Rigorous statistical analysis must be employed to provide confidence in survey

results.

- Selection of websites to be surveyed must be undertaken in a definitive and reproducible manner.

Unfortunately, the original AT&T survey does not possess many of these characteristics, causing considerable problems as outlined in the following sections.

2.2.1 The AT&T Survey Methodology

Through the use of the AT&T Privacy Bird P3P agent, a series of websites belonging to multiple lists were surveyed in May 2003 to determine if they possessed P3P policies. The retrieved policies were then analyzed to determine the types of data used by each website.

2.2.1.1 Site Selection

Cranor, Byers, and Kormann [36] chose to select a series of websites that were popular or located on popular indexes or lists instead of implementing a random sampling methodology (Table 2-1). Their rationale for this originates from a desire to analyze P3P

Table 2-1: Lists of Websites Used in the AT&T Survey

List	Source	Date	Description	Currently Available
PFF Most Popular	Popularity rankings by Neilson/NetRatings	Oct-01	Contains 85 of the 100 busiest websites	No
PFF Random	Popularity rankings by Neilson/NetRatings	Oct-01	302 sites selected at random from the top 7,821 websites	No
PFF Refined Random	Popularity rankings by Neilson/NetRatings	Oct-01	209 sites selected at random from the top 5,625 websites	No
Netscore top 500	Popularity rankings by comScore media metrix	Jul-02	Top 500 websites	No
Key Measures	Popularity rankings by comScore media metrix	Jul-02	Top 500 websites	No
Alexa Global 500	Popularity rankings	Feb-03	Top 500 websites rated by Alexa	Yes
Froogle	Web Crawl of froogle.com	Apr-03	Businesses indexed in Froogle.com	No
Yahooligans	WebCrawl of yahooligans.com	Apr-03	Children's websites indexed on yahooligans.com	No
FirstGov	Compiled from firstgov.gov	Apr-03	The US government websites indexed on firstgov.gov	Yes
News	Web Crawl of news.google.com	Apr-03	News websites indexed on news.google.com	No

from a user's perspective.

While we concur with this choice, there are several problems with the lists selected. Some lists (PFF Top 100, PFF Refined Random, and PFF Random) were constructed using rankings no longer publicly available; in addition, the authors of the lists arbitrarily removed "adult", "children", and "business to business" websites without providing a precise definition of what these terms mean or how their 'removal' policy was implemented. The Key Measures and Netscore Top 500 lists are also no longer available and the Netscore Top 500 list removed sites satisfying the criteria of "third party sites including advertising networks" where third party and advertising networks were not clearly defined. Further, the decision criteria for link selection in the web-crawler employed to gather the Froogle, Yahooligans, and News lists was not explained. This is important because a website may contain links that belong to a different category; for example news.google.com links to impages.google.com, which is not a news site, and may possess a different P3P policy.

2.2.1.2 Collection of Data

The AT&T survey used the P3P analysis engine of the AT&T Privacy Bird P3P agent [39] that was previously implemented by the survey's authors. This engine provides tools for XML analysis and the definition of privacy preferences using the APPEL rule-based language [42]. These abilities allowed for the following four types of data to be collected: information pertaining to XML structural correctness, P3P document validity, frequency of P3P tag occurrence, and a categorization of policies based on perceived safety, subjectively defined by the AT&T Privacy Bird creators.

The use of this engine introduces an implementation bias that is incompatible with the implementations of the official P3P analysis tools provided by the W3C. The engine records errors if omissions or mistakes are found in the document that leads to the policy being unusable. If the policy contains errors that are not critical, then no error will be reported [36]. For this reason, AT&T Privacy Bird may be able to analyze policies that would cause other agents to fail. In addition to errors, implementation bias is introduced through the categorization of policies using the predefined APPEL rule sets in the engine. The rule sets in question categorize websites based on a perceived level of risk applied to certain actions on individual data items. However, this level of risk is highly dependent on factors outside of P3P's scope such as situational context [43].

2.2.1.3 Analysis of Data

The majority of the data analysis discussed general P3P adoption by list and relative frequency of P3P tags by list. The analysis pertaining to the general adoption of P3P detailed both the number of sites implementing P3P as well as the number of sites whose P3P documents were found to be unusable. However, inferences made upon differences in the observed data are questionable due to the lack of statistical analysis. The analysis concentrating on relative frequency of P3P tags adds little value to the goals of our project and for this reason it is only mentioned to ensure a complete overview of the AT&T survey. Finally, it is noted that the survey did not report on the adoption of P3P compact policies except to state an overall adoption rate for all analyzed lists. No discussion was provided describing the extent of P3P compact policy adoption in individual website populations.

2.2.2 2005 Study Methodology

The 2005 study uses a site selection mechanism that is similar to the AT&T survey, but utilizes an official policy verification tool provided by the W3C and will extend and improve upon the analysis undertaken in the AT&T survey by using a greater number of business or E-Commerce lists, a greater number of websites surveyed, and rigorous statistical analysis. This survey methodology was implemented in both the February 2005 pilot study as well as our full November 2005 study.

2.2.2.1 Site Selection

Table 2-2 outlines the comparisons undertaken between the 2005 studies and the AT&T survey. The Alexa Global 500 [44]and First Gov [45] lists exist in both studies so no approximation is required. Alexa.com is a respected subsidiary of Amazon.com that collects traffic information that is then used by Amazon.com and other affiliated business partners to analyze Internet traffic patterns. The FirstGov list contains all United States federal and state websites, allowing for an analysis of P3P adoption in an environment where machine-readable privacy policies must be provided. The remainder of the lists in Table 2-2 requires approximations. Website rankings from Alexa.com are used to approximate the PFF lists as both are based upon popularity rankings. The ranking methods employed by Alexa.com and the PFF studies do differ slightly; the PFF lists estimate website popularity by the number of unique visitors per month, whereas the Alexa rankings are the geometric mean of both the number of unique users per day and

Table 2-2: Lists Comparable with the AT&T Survey

List	Source	Date collected	**Description**	Comparable list in AT&T survey
Alexa Global 500	Alexa.com	Feb 2005	Top 500 websites ranked on unique visitors and page views	Alexa Global 500
FirstGov	Firstgov.gov	Feb 2005	US government websites	First Gov
Alexa Top 100	Alexa.com	Nov 2005	Top 100 websites ranked on unique visitors and page views	PFF most popular
Alexa Top 7,821	Alexa.com	Nov 2005	Top 7,821 websites ranked on unique visitors and page views	PFF random
Alexa Top 5,625	Alexa.com	Nov 2005	Top 5,625 websites ranked on unique visitors and page views	PFF refined random
Top 400 Ecommerce	Internet retailer magazine	Nov 2005	Top 400 E-Commerce websites ranked on revenue	Froogle

the number of unique pages the user viewed per day for a specified time period. In addition, the exclusion of "Adult", "Children", and "Business to Business" websites from the PFF lists creates a further inconsistency between the two studies since no sites were excluded in the 2005 survey. The Froogle list is approximated using the Top 400 E-Commerce websites ranked by Internet Retailer Magazine [46]. To our knowledge, no other organization has amassed as large or as accurate a list as Internet Retailer Magazine. The Top 400 E-Commerce list is ranked on revenue generated from the online sale of physical items. While both of these lists are intended to target E-Commerce sites, it is unclear as to whether the Froogle list is entirely populated by E-Commerce sites; the web-crawler used to collect this list is never fully described. The comparison between the Froogle and Top 400 E-Commerce list is further complicated by the fact that the Top 400 E-Commerce list is ranked by online revenue (E-Commerce sales) while the Froogle list does not use this financial metric.

Table 2-3 describes the lists that will allow an extension of our results beyond the AT&T survey. The Top 300 E-Commerce list [47] was used in the February pilot survey because its replacement, the Top 400 E-Commerce list, had not yet been released. The

Table 2-3: Lists that will be Used in the Extended Analysis

List	Source	Date collected	Description
Top 300 E-Commerce	Internet retailer magazine	February 2005	Top 300 E-Commerce websites ranked on revenue for 2004
BBBOnline	BBBonline.org	February 2005	All websites/businesses that qualify for the BBBOnline reliability seal
Truste	Truste.org	February 2005	All websites/businesses that qualify for the Truste seal
Alexa Language Lists	Alexa.com	February 2005	20 languages, each containing 100 websites ranked on unique visitors and page views
Forbes 500	Forbes Magazine	February 2005	The Top 500 international companies ranked on revenue

BBBOnline list [48] provides a collection of business websites who have registered for the Better Business Bureau's web-reliability seal. A web-reliability seal indicates a trusted third party (BBBOnline) has certified that the business practices of the website in question meet the third party's standards of ethical business practice (including privacy protection). BBBOnline's parent, the Better Business Bureau, has operated as a third party recommendation organization in North America since 1912. The Truste list [49] targets a similar domain of business websites that have shown a previous interest in third-party certifications. Truste has operated as a non-profit third party recommendation service since 1997. Truste and BBBOnline are the two largest online recommendation organizations and the addition of these lists allows for analysis of websites who have shown a previous desire to implement technologies that inform users about a website's business practices. The inclusion of the Forbes 500 [50], which is generally considered the most accurate listing of large companies, allows for insight regarding P3P adoption in the largest companies in the world.

Finally, a major limitation of the AT&T survey is its bias towards P3P adoption in English-language websites. The addition of the Alexa Language lists [44] will allow us to explore P3P adoption in a broader linguistic and cultural context that to our knowledge has not been attempted in any previous survey. If it is found that P3P adoption is biased towards a particular culture or language, this problem must be remedied since it is unreasonable to expect all privacy conscious individuals to speak English.

While our lists cannot be considered representative of the general Internet, our lists do represent sectors of the Internet where important privacy concerns exist. Popular websites (Alexa Global 500, Alexa Language Lists) have the opportunity to influence online privacy since their actions affect large numbers of users. Business websites (Forbes, Top 300/400 E-Commerce, BBBOnline, and Truste) may implement user profiling, data aggregation, and differential pricing. Finally, government websites (FirstGov) often utilize highly sensitive information (Social Security Numbers and health/financial information), which if mishandled can have serious consequences to a user.

2.2.2.2 Collection of Data

Instead of using the AT&T Privacy Bird engine, our survey used the official P3P validator tool [51], provided by the W3C, to harvest P3P policies and extract relevant information. This choice is due to the lack of strict adherence to the P3P protocol by the AT&T Privacy Bird engine regarding handling policies with errors [36]. While this lack of adherence may be desirable in a fault tolerant P3P agent, it is not advantageous when one is attempting to scientifically analyze P3P adoption.

2.2.2.3 Analysis of Data

For each of the points of interest between the two studies, we formally investigate the change in the representative dichotomous variables and report on the relative significance between each of these relationships. We decided that a common approach to this analysis was preferable, and during the initial design of the survey, we had limited insight into the likely values that each of the dichotomous variables would possess – and could not rule out a series of low frequency values. Given these assumptions, it was believed that an application of Fisher's Exact probability test was more appropriate than a Chi-squared test (with or without Yates' correction), which has difficulties accommodating low frequency values and does not naturally lend itself to directional formations. Having said this, we note that in general, Fisher's Exact Probability test possesses a lower statistical power than the corresponding Chi-squared test and hence our analysis can be viewed as conservative in this respect [52]. Finally, due to what we see as conflicting causal evidence on adoption drivers for P3P, all statistical tests are formulated as two-tailed, non-directional, tests and α is set at the traditional 0.05 level.

Due to our desire to investigate P3P adoption in a general sense, analysis of dichotomous variables must be expanded beyond individual lists. This analysis is complicated by the existence of sample sizes varying from 300 to 20263 websites; a simple amalgamation of unique websites would bias the results towards the large lists. We decided to employ resampling techniques to create representative, amalgamated (from multiple lists) distributions of the union of individual lists by randomly sampling each list 10000 times. This process results in the construction of a Gaussian distributed description of the entire population [53]. Each population was amassed through the selection of an equal number of websites from each list, chosen randomly and without replacement from their respective list. Due to the limited guidance in the relevant literature regarding how large a sample should be, we arbitrarily decided to collect 200 websites from each list. This choice allows for a random selection from all lists to occur, and when we tested our results against other sample sizes we found little difference in results. Duplicated websites in the lists were removed once the representative populations were created; this choice ensured that the created populations are realistic (no duplicates), and that lists whose websites were selected last were not over-represented.

The distributions representing the dichotomous variables contained in these populations are analyzed through one-way ANOVA ($\alpha = 0.05$) and Cohen's d effect size (including bias corrected equivalents) [54] methods. The use of statistical significance tests alone is unreasonable due to the one-way ANOVA's sensitivity to sample size, which conflicts with the need for large populations in Resampling techniques [55]. Cohen's d effect size allows for an analysis independent of sample size indicating the magnitude of the experimental effect. Since there is no method for calculating Cohen's d effect sizes directly from the contingency tables used during the application of Fisher's Exact test, we will first calculate an odds ratio effect size [56] which are then converted into a Cohen's d effect size using the method described by Hasselblad and Hedges [57]. Our analysis will follow Cohen's suggestion that d ± 0.2 indicates a small effect, d ± 0.5 is a medium effect, and d ± 0.8 is a large effect [54].

2.3 Comparison of the AT&T and 2005 Studies

The comparison between our November 2005 study and the AT&T survey will be completed in two parts. The analysis is divided due to the list approximations that led to a decreased confidence in results. Table 2-4 examines full policy adoption in the lists whose domains can be considered the same with a high degree of confidence. The high degree of confidence stems from both the sources being identical and the lack of any statistically significant difference in availability. The results indicate a disparity in full policy adoption between U.S. government websites and the most popular websites on the internet. Government websites have seen a statistically significant ($p < 0.0001$), large ($d = 1.6537$) increase in full policy adoption. The increase in adoption is likely a result of the enacting of the E-Governance act of the United States [58]. This act requires government websites to provide their privacy policies in a machine interoperable manner. In contrast, no statistical evidence exists indicating a change in adoption for the Alexa Top 500 websites ($p < 0.3548$).

Table 2-4: P3P Adoption in Lists That Do Not Require An Approximation

	Firstgov		Alexa Global 500	
Date	July-03	Nov-05	July-03	Nov-05
Total Sites	344	366	500	500
Total Accessible	338	359	495	489
Availability P-value	1.0000		0.2064	
Total P3P enabled	7	107	92	79
P3P enabled P-value	0.0001		0.3548	

The analysis of Table 2-5 assumes that the samples (May 03, Nov 05) are drawn from the same underlying distribution and can be considered to be a representative sample of the distribution. Given the AT&T survey either chose to exclude certain "categories" of data or didn't provide suitable description of their collection techniques, it is debatable as to the validity of the assumption as this is likely to have skewed the sample. Hence, we urge caution when interpreting the results from Table 2-5. A further concern is with regard to website accessibility between competing lists skewing the results. This is analyzed in line 2 of Table 2-5. As can be seen, statistically significant differences exist in the number of accessible sites for the comparisons between PFF refined random/Alexa 5,625 ($p < 0.0001$) and PFF random/Alexa 7,821 ($p < 0.0004$); however, the associations are unlikely to have a substantial effect as they exhibit effect sizes of $d = 0.2591$ and $d =$

Table 2-5: Analysis of Lists Requiring an Approximation for Comparison Between 2003 and 2005

	Froogle	Top 400 E-Commerce	PFF Top 100	Alexa Top 100	PFF Refined Random	Alexa Top 5,625	PFF Random	Alexa Top 7,821
Total Sites	1017	400	85	100	209	5,625	302	7821
Total Accessible sites	1010	394	84	100	195	5519	286	7672
Accessible sites p-value	0.2111		1.0000		0.0001		0.0004	
Total sites P3P enabled	133	121	26	17	29	540	35	669
P3P enabled p-value	0.0001		0.0353		0.0275		0.00438	

0.2235 respectively indicating a relatively small effect.

If one assumes the distributions are suitably similar, then there is evidence indicating that all four groups of websites experienced a statistically significant change in full policy adoption (line 5). Further ad-hoc analysis shows that an increase in full policy adoption occurred for E-Commerce sites (Froogle/Top 400 E-Commerce), whereas a decrease has occurred in the remaining groups. We are unable to provide any explanation for the divergence of these results.

2.4 Comparison of the February and November 2005 Studies

Table 2-6 summarizes how P3P adoption has changed between Feb and Nov 2005. Our confidence in the results is high due to the lists being relatively independent (Table 2-7), with most of the overlaps being sufficiently under 10%. In addition, no statistically significant differences in website availability exist except for the BBBOnline and

Table 2-6: Comparison of February 2005 and November 2005

	Alexa Global 500		BBBOnline		FirstGov		Forbes 500		Top 300 E-Commerce		Truste	
	Feb	Nov	Feb	Nov	Feb	Nov	Feb	Nov	Feb	Nov	Feb	Nov
Total Sites	500	500	20263	20263	366	366	500	500	300	300	1317	1317
Total Accessible Sites	480	489	19322	19059	366	359	493	496	298	293	1277	1286
Accessible p-value	0.1433		0.0001		0.0152		0.5466		0.1768		0.3359	
Total sites with full policy	77	79	844	980	81	107	24	27	87	87	192	202
P3P enabled p-value	1.0000		0.0004		0.0220		0.7740		0.9282		0.6614	
Total sites with compact policy	60	68	770	1045	11	17	12	15	89	82	149	149
Total compact policy p-value	0.5694		0.0001		0.2517		0.6971		0.6505		0.9510	
Total sites with full and compact policy	49	46	458	630	9	14	7	8	59	49	91	94
P3P compact policy p-value	0.7460		0.0001		0.2956		1.0000		0.3404		0.8789	

Firstgov lists. In the case of BBBOnline, while the difference is statistically significant, the association is unlikely to have a substantial effect as it only exhibits a small effect size of d=0.1446. To avoid problems of small cell frequencies when calculating the odds ratio effect size for the accessibility of websites for the Firstgov list, we will add 0.5 to the observed frequencies in accordance with accepted practice [59]. The resulting large (d = 1.5036) operational effect detected in the FirstGov list suggests a major changed occurred, and while 98.1% of the sampled websites were still able to be contacted, caution is urged when interpreting these results.

The FirstGov ($p < 0.0220$) and BBBOnline ($p < 0.0004$) lists were the only lists that experienced a statistically significant difference in adoption of full P3P policies. Through ad-hoc analysis, full policy adoption was found to be increasing for both FristGov (likely due to the enacting of the E-Governance Act) and BBBOnline. We know of no rational for the differences between BBBOnline and the other lists. BBBOnline was also the only list that experienced a statistically significant change in P3P compact policy adoption.

It is surprising that many websites adopt full P3P policies without compact policies since rational organizations would only adopt a technology when suitable incentives exist. Given the lack of usable P3P user agents or laws requiring P3P adoption for private organizations, the most significant incentive to adopt P3P appears to be the Internet Explorer cookie blocking feature. Our results however suggest that some other motivation appears to exist. However, due to limitations in our dataset, we are unable definitively ascertain what these motivations may be. The results of Table 6 also suggest that the P3P protocol is often abused since many websites using compact policies do not post a corresponding full policy. This result brings into question the inclusion of P3P

Table 2-7: Overlap Between Lists in November 2005

	Alexa Global 500	FirstGov	Top 300 E-Commerce	Forbes 500	Truste	BBBOnline
Alexa Global 500	-	75.00% (3) [4] 0.80%	38.89% (7) [18] 3.60%	38.46% (5) [13] 0.60%	29.41% (20) [68] 13.60%	33.33% (6) [18] 3.60%
FirstGov	75.00% (3) [4] 1.09%	-	0.00% (0) [0] 0.00%	0.00% (0) [0] 0.00%	0.00% (0) [0] 0.00%	0.00% (0) [0] 0.00%
Top 300 E-Commerce	38.89% (7) [18] 6.00%	0.00% (0) [0] 0.00%	-	23.08% (3) [13] 4.33%	23.53% (8) [34] 11.33%	56.10% (23) [41] 13.67%
Forbes 500	38.46% (5) [13] 2.60%	0.00% (0) [0] 0.00%	23.08% (3) [13] 2.60%	-	37.50% (3) [8] 1.60%	28.57% (2) [7] 1.40%
Truste	29.41% (20) [68] 5.16%	0.00% (0) [0] 0.00%	23.53% (8) [34] 2.58%	37.50% (3) [8] 0.61%	-	19.51% (32) [164] 12.45%
BBBOnline	33.33% (6) [18] 0.09%	0.00% (0) [0] 0.00%	56.10% (23) [41] 0.20%	28.57% (2) [7] 0.03%	19.51% (32) [164] 0.81%	-

Ex. The overlap between Alexa Global 500 and FirstGov consists of 4 sites. Of these 4 sites, 3 are P3P enabled, leading to a P3P adoption rate of 75% in the overlap. Additionally, Four websites exist in both Alexa Global 500 and FirstGov, thus the percentage of the FirstGov list that overlap with Alexa Global 500 is 4/366 = 1.09% and the percentage of Alexa Global 500 websites that overlap with FirstGov is 4/500 = 0.80%.

compact policies in the P3P specification since the only method available to determine if a website is adhering to the P3P protocol is to attempt to request the full P3P policy, thereby negating any performance improvements.

Table 2-7 indicates another unexpected result suggesting that when a website exists in more than two lists from Table 2-6, it is often more likely to possess a P3P document. Unfortunately, analysis beyond this informal statement would be questionable due to the generally small magnitude of the overlaps. This analysis is only provided for completeness and to point out a peculiar effect that we are currently unable to explain.

Since no generalizations of P3P adoption can be determined form the results of Tables 2-6 and 2-7 , a resampling based analysis will now be conducted (Table 2-8). In this analysis, four groups were developed from the lists analyzed in Table 2-6. These groups are based upon general characteristics inherent to the websites belonging to the lists and due to space constraints, only full P3P policies will be analyzed. We urge caution when interpreting our groups as true indicators of their respective Internet fields. It is possible that our groups lack important subsets of Internet sites. In addition, our groups may suffer from an overemphasis of lists such as FirstGov that could have unique P3P adoption drivers (E-Governance Act). The 'All Lists' group is an amalgamation of all six lists contained in Table 2-6. The results for this group represent general P3P adoption over all surveyed lists. The 'Non-Legislated' group is comprised of all websites not required by law to adopt P3P or a similar substitute (All lists except FirstGov). This group will provide insight as to whether website administrators choose to adopt P3P on their own accord. The 'Business' group is comprised of all websites belonging to Forbes 500, Top 300 E-Commerce, BBBOnline, and Truste. These lists have a business focus where the website's functionality allows for the disseminating of company information and/or directly transacting with customers. For analysis of sites that are explicitly transaction based, refer to our analysis in Table 2-6 regarding Top 300 E-Commerce. Our final group is the 'Seal Adopting' group (BBBOnline and Truste).

Table 2-8: General P3P Adoption in Various List Groupings Using Resampling Techniques

		All Lists		Non-legislated		Business		Seal adopting	
		Feb 05	Nov 05	Feb 05	Nov 05	Feb 05	Nov 05	Feb 05	Nov 05
Number of generated populations		10000		10000		10000		10000	
Mean number of P3P policies		175.46	194.8	131.44	135.13	103.75	107.1	37.98	40.21
Std deviation		8.46	8.96	7.66	7.80	6.80	7.05	5.40	5.52
Std errors		0.085	0.090	0.077	0.078	0.068	0.071	0.054	0.055
F-value for change in number of P3P policies		24640.478		1137.956		1171.800		832.363	
p-value for change in number of P3P policies		0.0001		0.0001		0.0001		0.0001	
Cohen's d effect size for change in number of P3P policies		2.22		0.48		0.41		0.48	
Std error of the effect size estimate		0.02		0.01		0.01		0.01	
Confidence interval	Lower	2.18		0.45		0.46		0.38	
	Upper	2.25		0.51		0.51		0.44	

The results of Table 2-8 indicate that the only group which experienced a significant operational alteration (d > 0.8, large) was the 'All Lists' group; this group experience a significant increase in adoption between the control (Feb 2005) and treatment (November 2005) groups. The confidence intervals indicate that this large effect contains very little random variation due to experimental effects or sample limitations and hence can be considered a highly reliable figure. However, when we look at the non-legislative group, we see the operational alteration decreased substantially to a small/medium (0.2 < d < 0.5) effect indicating that much of the effect in the 'All Lists' group is attributable to the websites within the FirstGov list. The other groups experience effect sizes broadly inline with this 'Non-Legislative' group and hence we might hypothesize that outside of the (U.S.) governmental sector, P3P implementation rates were relatively stable throughout the study. Confidence intervals for these groups indicate that these results are reasonably robust to experimental variation.

2.5 P3P Document Errors/Maintenance Between February 2005 and November 2005

While completing the analysis in Sections 2.3 and 2.4, a large number of errors were encountered in all lists (Table 2-9). Table 2-9 only reports errors that are strict violations of the P3P XML schema or basic XML document structure; this does not include inconsistencies between P3P policies and their human readable counterparts. We will use the P3P validator which is an open source application provided by the W3C and is the official P3P document verification tool for identifying violations of the P3P schema and XML document syntax. These errors can have a significant effect on the usability of these documents since some errors may render the document unusable. For instance, the semantic meaning of a document can become indeterminate if a start-tag (<myTag>) exists without a corresponding end-tag (</myTag>) [60]. The analysis of inconsistencies between P3P policies and their human readable counterparts is not included since the analysis is a complex task requiring extensive knowledge of applicable legislation, case law, and expert opinion, which is beyond the scope of this paper. While Table 2-9 only indicates the prevalence of structural errors, no statistically significant changes were

Table 2-9: Percentage of P3P Policies Containing Structural Errors

		Feb-05	Nov-05	p-value
Alexa Global 500	Full	20.78%	22.78%	0.8470
	Compact	40.00%	35.29%	0.5894
BBBOnline	Full	60.55%	69.29%	0.0001
	Compact	13.51%	10.24%	0.0378
FirstGov	Full	9.88%	13.08%	0.6478
	Compact	27.27%	17.64%	0.6525
Forbes 500	Full	29.17%	25.93%	1.0000
	Compact	8.33%	7.14%	1.0000
Top 300 E-Commerce	Full	27.59%	33.33%	0.5102
	Compact	12.36%	12.20%	1.0000
Truste	Full	26.56%	29.21%	0.5760
	Compact	33.56%	30.87%	0.7101

detected except in the case of the BBBOnline group. Further ad-hoc analysis indicates that the BBBOnline group experienced an increase in the error rates for full policies but a decrease in the error rates in compact policies between Feb and Nov 2005. For full policies, the operational effect of the statistically significant difference was small/medium (d = 0.212). For compact policies, the statistically significant difference was found to have a small operational effect (d = 0.173) respectively, suggesting little change occurred over the time period.

It is interesting to note that the prevalence of errors can vary widely between full and compact policies. It is especially surprising that P3P compact policies were more likely to be in error than full policies for the Alexa Global 500, First Gov, and Truste groups. Compact policies are simpler (a string of text) and do not need to conform to stringent XML formatting rules. We are currently at a loss to explain these results.

Given the relatively large number of policies that are invalid (Table 2-9), a more detailed, yet rigorous analysis of these errors is conducted. Table 2-11 describes how P3P full and compact policies are being adopted and whether they contain errors. Table 2-10 describes the extent that non-corrective maintenance is being undertaken, and finally, Table 2-12 describes the extent that corrective maintenance is undertaken. The largest improvement in full P3P policies occurred in the Forbes 500 list where 1 out of 7 invalid policies were improved (Table 2-12) and from this result, it would appear that few invalid full or compact P3P policies are being corrected. In addition to the lack of corrective maintenance, we detected very little non-corrective maintenance of P3P policies (Table 2-10). A P3P policy (full or compact) is deemed to have undergone non- corrective

Table 2-11: Total Number of Sites with Valid, Invalid, and No P3P Full/Compact Policies as of Feb 2005

Status of Full Policy	Status of Compact Policy	Alexa Global 500	BBBOnline	FirstGov	Forbes 500	Top 300 E-Commerce	Truste
Valid	Valid	20 (4.2%)	112 (0.6%)	7 (1.9%)	5 (1.0%)	39 (13.1%)	41 (3.2%)
Valid	Invalid	17 (3.5%)	8 (0.0%)	1 (0.3%)	1 (0.2%)	2 (0.7%)	29 (2.3%)
Valid	None	24 (5.0%)	213 (1.1%)	65 (17.8%)	11 (2.2%)	22 (7.4%)	71 (5.5%)
Invalid	Valid	12 (2.5%)	327 (1.7%)	0 (0.0%)	1 (0.2%)	18 (6.0%)	21 (1.6%)
Invalid	Invalid	0 (0.0%)	11 (0.0%)	1 (0.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Invalid	None	4 (0.83%)	173 (0.9%)	7 (1.9%)	6 (1.2%)	6 (2.0%)	30 (2.3%)
None	Valid	14 (2.9%)	227 (1.2%)	1 (0.3%)	5 (1.0%)	21 (7.0%)	37 (2.9%)
None	Invalid	7 (1.5%)	85 (0.4%)	1 (0.3%)	0 (0.0%)	9 (3.1%)	21 (1.6%)
None	None	382 (79.6%)	18166 (94.0%)	283 (77.3%)	464 (0.94%)	181 (60.7%)	1027 (80.4%)

maintenance if any information elements in the document changed while remaining valid in both studies. The low frequency of non-corrective maintenance precludes any application of statistical tests in identifying major changes. Through a visual inspection

Table 2-10: Number of Sites Performing Non-Corrective Maintenance of P3P Policies

	Alexa Global 500	BBBOnline	FirstGov	Forbes 500	Top 300 E-Commerce	Truste
Total number of P3P enabled sites in both Feb and Nov with valid P3P policies	54	261	66	15	53	131
Number of valid full or compact policies that were modified and remained valid.	0	5	5	0	1	2

of the modifications, it was found that many of the changes involved types of data collected, purpose for collection, and length of retention information. It would appear reasonable to expect such changes as business practices within organizations change.

In order to generalize our findings, Resampling techniques using the groups defined in Table 2-8 were again employed to analyze the occurrence of valid (Table 2-13) and invalid (Table 2-14) full policies. Table 2-13 indicates that the 'All lists' group experienced a large ($d > 0.8$) increase in valid full policy adoption, which can be stated with a high degree of confidence (lower confidence interval = 1.29). In contrast, all other groups either experienced small ($d > -0.2$) ('Seal Adopting' websites) or small/medium ($-0.5 < d < -0.2$) decreases in valid full policy adoption! For example, 'Non-legislated' ($d = -0.29$) and 'Business' ($d = -0.35$) groups experienced small to medium decreases. Again, these results can be stated with a 95% certainty due to the small range of the confidence interval. We are at a loss to explain why full policy adoption is so low for 'seal adopting' sites when they have shown a previous inclination to adopt technologies with a similar purpose. We can only speculate that this may be due to a lack of requests for such documents from website patrons.

Table 2-12: Number of Sites Modifying, Adopting, or Dropping Their P3P Enabled Status

	Alexa Global 500	BBBOnline	FirstGov	Forbes 500	Top 300 E-commerce	Truste
Available in both Feb and Nov 05	475	18808	359	491	293	1262
Full Policies						
Invalidated their policy (Valid -> Invalid)	1 (0.2%)	4 (0.0%)	4 (1.1%)	0 (0.0%)	3 (1.0%)	1 (0.0%)
Fixed their policy (Invalid -> Valid)	1 (0.2%)	1 (0.0%)	1 (0.3%)	1 (0.2%)	1 (0.3%)	3 (0.2%)
Dropped a valid policy (Valid -> Dropped)	6 (1.3%)	58 (0.3%)	1 (0.3%)	2 (0.4%)	7 (2.4%)	9 (0.7%)
Dropped an invalid policy (Invalid -> Dropped)	2 (0.4%)	44 (0.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (0.2%)
Created a valid policy (No P3P -> Valid)	6 (1.3%)	39 (0.2%)	26 (7.2%)	4 (0.8%)	4 (1.4%)	9 (0.7%)
Created an invalid policy (No P3P -> Invalid)	4 (0.8%)	222 (1.2%)	4 (1.1%)	1 (0.2%)	4 (1.4%)	8 (0.6%)
Compact Policies						
Invalidated their policy (Valid -> Invalid)	0 (0.0%)	2 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.0%)	0 (0.0%)
Fixed their policy (Invalid -> Valid)	1 (0.2%)	5 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (0.2%)
Dropped a valid policy (Valid -> Dropped)	6 (1.3%)	42 (0.2%)	0 (0.0%)	0 (0.0%)	9 (3.1%)	6 (0.5%)
Dropped an invalid policy (Invalid -> Dropped)	1 (0.2%)	6 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.3%)	3 (0.2%)
Created a valid policy (No P3P -> Valid)	10 (2.1%)	293 (1.6%)	6 (1.7%)	2 (0.4%)	2 (0.7%)	4 (0.3%)
Created an invalid policy (No P3P -> Invalid)	4 (0.8%)	7 (0.0%)	0 (0.0%)	1 (0.2%)	5 (1.7%)	2 (0.2%)

Table 2-13: Analysis of Valid P3P Policy Adoption in Various Internet Domains Using Resampling Techniques

		All Lists		Non-legislated		Business		Seal adopting	
		Feb 05	Nov 05	Feb 05	Nov 05	Feb 05	Nov 05	Feb 05	Nov 05
Number of generated populations		10000		10000		10000		10000	
Mean number of Valid P3P policies		133.89	144.08	94.24	92.35	72.73	70.71	25.44	25.32
Standard Deviation		7.51	7.87	6.54	6.58	5.72	5.81	4.45	4.43
Standard Errors		0.075	0.079	0.065	0.066	0.057	0.058	0.044	0.044
F-value for change in number of valid P3P policies		8777.429		416.116		615.732		3.470	
p-value for change in number of valid P3P policies		0.0001		0.0001		0.0001		0.063	
Cohen's d effect size for change in number of valid P3P policies		1.32		-0.29		-0.35		-0.03	
Standard error of the effect size estimate		0.02		0.01		0.01		0.01	
Confidence interval	Lower	1.29		-0.32		-0.38		-0.05	
	Upper	1.36		-0.26		-0.32		-0.00	

Table 2-14: Analysis of Invalid P3P Policy Adoption in Various Internet Domains Using Resampling Techniques

		All Lists		Non-legislated		Business		Seal adopting	
		Feb 05	Nov 05	Feb 05	Nov 05	Feb 05	Nov 05	Feb 05	Nov 05
Number of generated populations		10000		10000		10000		10000	
Mean number of Invalid P3P policies		41.57	50.72	37.20	42.78	31.02	36.39	12.54	14.89
Standard Deviation		4.71	5.05	4.51	4.73	4.18	4.48	3.30	3.58
Standard Errors		0.047	0.051	0.045	0.047	0.042	0.045	0.033	0.036
F-value for change in number of invalid P3P policies		17479.022		7285.113		7706.579		2313.170	
p-value for change in number of invalid P3P policies		0.0001		0.0001		0.0001		0.0001	
Cohen's d effect size for change in number of valid P3P policies		1.87		1.21		1.20		0.68	
Std error of the effect size estimate		0.02		0.02		0.02		0.01	
Confidence interval	Lower	1.84		1.18		1.17		0.65	
	Upper	1.91		1.24		1.23		0.71	

The results of Table 2-14 (invalid full policy adoption) indicate that a medium/large (0.5 < d < 0.8) increase of invalid full policies occurred in 'Seal Adopting' websites (d = 0.68). However, in the cases of 'All Lists', 'Non-Legislated', and 'Business' groups, large (d > 0.8) increases in invalid full policy adoption were found with a high degree of confidence. We are at a loss to explain these error rates due to the free and publicly available W3C P3P validator tool, which analyzes P3P policies for structural errors. This effect has some interesting parallels with recent surveys conducted by Tien *et al.* [61] and Huynh and Miller [62] who investigate defect rates found on web-servers. Both surveys show that the defects are dominated by 404 errors, which are primarily missing hypertext links. Again, these error types are completely preventable as "link checkers", which can automatically discover these errors, are widely available. Again, both surveys were unable to provide a causal explanation as to why these defects remain uncorrected. The results of Table 2-13 and Table 2-14 indicate that the increases in P3P adoption that were observed in Table 2-6 were primarily a result of invalid P3P policy adoption, with exception of the sites within the FirstGov list. We in fact have found a decrease in the deployment of valid P3P policies in all groups that did not include the FirstGov list!

2.6 P3P Adoption Across Languages and Jurisdictions

The Alexa language lists provide an opportunity to determine if full policy adoption is highly influenced by changes in language or jurisdiction. Until recently, the P3P specification was only available in English, potentially limiting P3P adoption. Additionally, since P3P relies upon existing legislation for enforcement and the actions websites can undertake may be limited by such legislation, it would appear reasonable to expect that P3P usage should vary between jurisdictions.

Figure 2-2 describes how differences in full policy adoption exist between the most popular 100 websites for 20 languages as of November 2005. Statistically significant differences exist between English-language lists and all other language lists except Dutch and French. Our confidence in these results is enhanced as there is no statistical evidence indicating changes in P3P adoption in any of the lists between the Feb. and Nov. 2005 studies.

These observed differences in P3P adoption could also conceivably arise due to different cultural or legal influences which could motivate website operators to more readily adopt P3P than others. In order to undertake such analysis, the nation each website in the Alexa language list is hosted from was identified through reverse IP lookups using the Linux 'host' program. Once retrieved, the IP addresses were compared to a database purchased from IP2Location [63] which maps IP addresses to a particular nation. IP2Location states that their accuracy is above 95%. Through this method, it was determined that the websites contained in the Alexa language lists originated from 49 nations. Some nations, such as the U. S., host websites from a variety of languages. For instance, 84 English, 47 Arabic, 26 Greek, and 26 Spanish websites are hosted from the United States. Websites from these languages constitute 60% of all websites hosted from the United States. Additionally, since many nations have few websites being hosted from them, a Pareto analysis with a cutoff of 90% was used to reduce the number of nations under analysis resulting in nations with fewer than 30 websites being discarded. The results of Table 2-15 indicate that P3P adoption varies between jurisdictions and while this analysis cannot exclude language as an explanatory factor in P3P adoption, it does suggest that

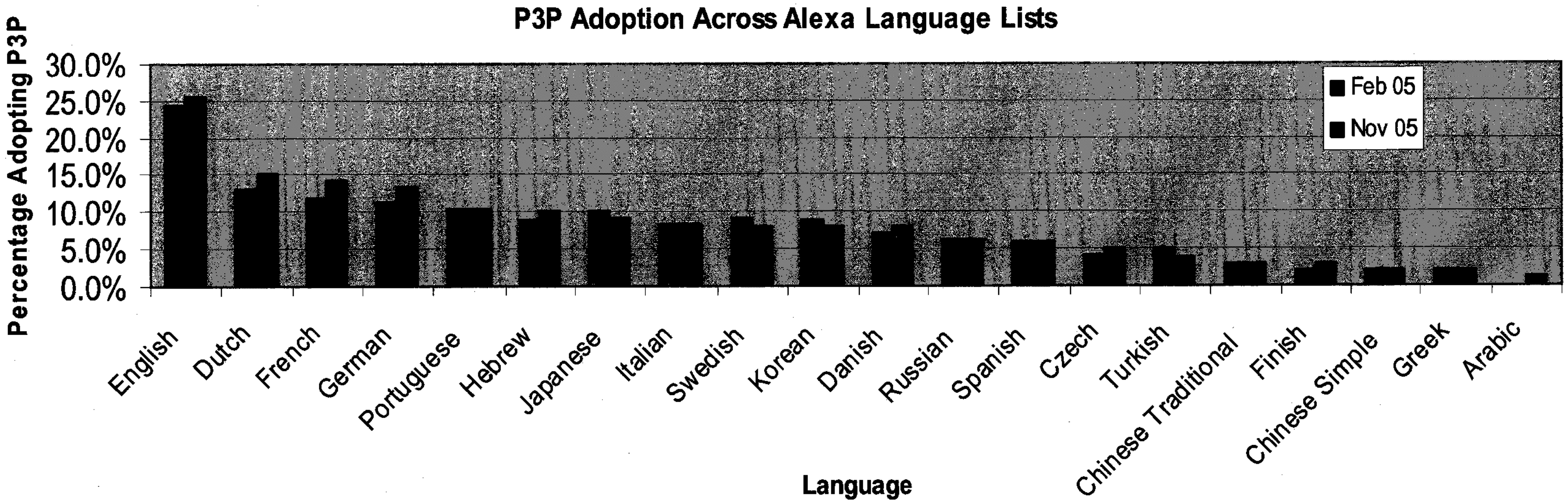

Figure 2-2: P3P Adoption by Language

Table 2-15: Analysis Of P3P Policy Adoption By Country

	Number of sites available		Number adopting p3p		Percentage adopting P3P	
	Feb	Nov	Feb	Nov	Feb	Nov
United States	297	298	34	38	11.4%	12.8%
Sweden	102	101	10	9	9.8%	8.9%
South Korea	94	93	8	7	8.5%	7.5%
Israel	90	90	6	7	6.7%	7.8%
Czech Republic	88	87	4	5	4.5%	5.7%
Finland	88	87	2	3	2.3%	3.4%
Japan	88	88	9	8	10.2%	9.1%
China	85	94	0	0	0.0%	0.0%
Denmark	83	82	6	7	7.2%	8.5%
Turkey	83	83	4	3	4.8%	3.6%
Russian Federation	79	79	5	5	6.3%	6.3%
Germany	77	77	10	12	13.0%	15.6%
Italy	75	76	5	4	6.7%	5.3%
Hong Kong	65	64	1	1	1.5%	1.6%
France	63	64	6	7	9.5%	10.9%
Greece	61	60	2	2	3.3%	3.3%
Canada	57	57	5	6	8.8%	10.5%
Brazil	57	58	4	5	7.0%	8.6%
Spain	34	34	2	2	5.9%	5.9%
United Kingdom	32	31	11	10	34.4%	32.3%

P3P is not well suited to particular legal or cultural domains since its adoption ranges from 32.3% in the United Kingdom to 0.00% in China. These results suggest that there may be a need for the application of culturally sensitive design practices [64] in future modifications of the P3P protocol.

2.7 P3P Adoption In the Internets Most Popular 100,000 Websites

After the pilot survey was completed, our results indicated that P3P adoption might be biased towards popular websites (the AT&T survey found P3P adoption to increase with popularity). In addition, we were curious to see if the observed error rates continued in the general Internet population. As of October of 2005, Alexa provided their rankings of the top 100,000 websites on the Internet. The addition of this list allows us to answer these questions regarding P3P adoption and errors. Figures 2 and 3 represent a centile-by-centile analysis of P3P full and compact policy adoption in the Alexa 100,000 list. Each centile is composed of 1000 websites grouped by rank. Caution is urged when interpreting the results of Figure 2-3 and Figure 2-4. A correct portrayal of the results would require a bar or scatter plot graph since the values derived from the centiles are discrete events. We chose to portray the graph as a line graph to ensure readability for the reader.

Analysis of Figure 2-2 indicates that P3P full policy adoption is correlated with website popularity with a non-linear relationship. P3P adoption rates start at approximately 15% for the most popular sites. Adoption then drops to a relatively consistent rate of about 3%

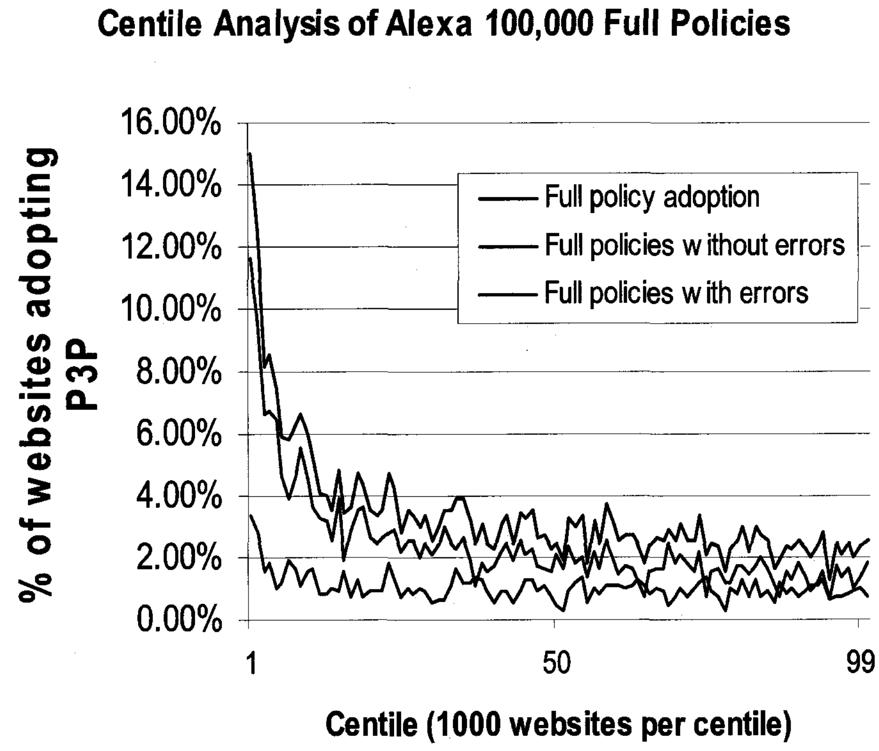

Figure 2-3: Full Policy Adoption for Alexa 100,000 Websites

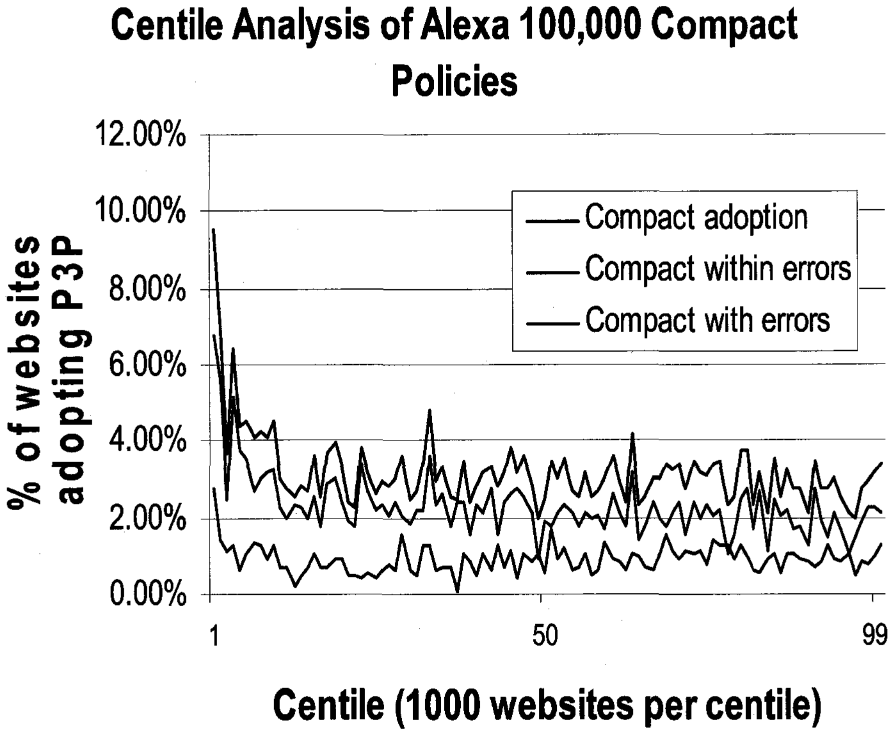

Figure 2-4: Compact Policy Adoption for Alexa 100,000 Websites

for less popular websites. Errors do not appear to follow the same non-linear relationship of general P3P adoption, but instead follow a relatively consistent rate of about 1% of websites.

Figure 2-4 describes the adoption of P3P compact policies. In general, the adoption curve of P3P compact policies follows the adoption of P3P full policies with the only major deviation occurring in the most popular websites on the Internet. This divergence in P3P full/compact policy adoption suggests that the motivation to adopt P3P may be different between the most popular websites on the Internet and the remainder.

2.8 Conclusion

Our results paint a sometimes contradictory and concerning image of P3P adoption. Evidence exists indicating that specific Internet domains have seen large increases in P3P adoption (FirstGov, Froogle/Top 400 E-Commerce) since the 2003 AT&T survey. However, only FistGov continues to show increasing adoption of valid P3P policies between February 2005 and November 2005. While this increase is a positive indicator for P3P, one has to wonder why more U.S. government websites have not adopted P3P give that they have had more than three years to adopt P3P since the enacting of the E-Governance Act. To our knowledge there is no other suitable substitute that would satisfy the requirements of the E-Governance act. It would appear reasonable to argue that P3P adoption would increase if P3P had been explicitly named in the E-Governance Act. Our dataset is however insufficient to allow us to comment on the effects of naming such a technology in legislation.

Our results also bring into question whether P3P can exist in a self-regulatory environment. Analysis of Non-Legislated websites indicates a decrease in valid P3P policies and an increase of invalid P3P policies. When one takes into account the lack of corrective maintenance of invalid P3P full and compact policies, it appears that companies have little incentive to provide quality documents in a self-regulated environment. This is especially surprising in the case of P3P compact policies since it would appear to be in the organizations best interest to ensure third party cookies are not blocked. The only domain that experienced an increase in valid full P3P policies was the FirstGov list, which is a legislated domain.

The lack of observed maintenance should also bring into question the information contained in P3P full and compact documents. There is little evidence indicating that websites ever fix their erroneous polices. In addition, there appears to be little or no effort to update P3P full and compact policies over a 9-month time span. It would seem unreasonable to think that business practices change so infrequently for Internet companies.

This lack of valid P3P adoption and maintenance suggests that if P3P is to be improved as a protocol, an analysis needs to be undertaken that describes why websites adopt P3P. Beatty et al. [65] have recently attempted to undertake such an analysis using Roger's technology adoption factors. In this article, the authors conclude that P3P user agents are have poor trialibility, observability, and are incompatible with many users past experiences. The results of this survey suggest that this analysis needs to be extended upon to include adoption factors from an organizational viewpoint.

The usability of P3P also appears questionable. AT&T Privacy Bird users find the system to be of questionable value given the current adoption rates. Our results indicate that this is unlikely to change since the only identified growth in usable P3P policies occurred in a legislated domain. To our knowledge the only domain of websites that are legislated to provide machine readable privacy policies are those belonging to the U.S. federal and state governments. Without a significant increase in P3P adoption, the usability of these agents from an end-user perspective appears questionable at best.

P3P 1.1 may have the potential to remedy some of these issues. Early drafts provide user agent guidelines, standardized definitions of XML tags, and XML tags for additional types of information. While these improvements are certainly necessary, they may be insufficient. For instance, AT&T Privacy Bird satisfies many of the user interface guidelines and it is unclear if any of the changes will motivate websites to post P3P policies. Additionally, the ability to represent additional data types may increase the complexity of P3P policies; given the high error rates for P3P 1.0 policies, this might not be beneficial.

This research is limited by its concentration only upon the server side adoption of the P3P protocol. While information describing the extent that users are requesting these policies and the agents they use would undoubtly provide insight, such an analysis is beyond the scope of this paper and would require a large sample of Internet users covering large cultural and geographic distances.

In conclusion, our results appear to indicate that P3P is the Internet's privacy standard in name only, since the vast majority of Internet websites lack valid P3P policies. At this point in time, we would have to conclude that P3P offers little assistance to the user. However, if legislation similar to the E-Governance act requiring a machine-readable form of the privacy policy were to be enacted over a broad section of the Internet, adoption of valid P3P policies might increase. If this occurs, P3P agent usefulness may improve.

Future work we plan to undertake in this area includes a more comprehensive analysis of errors in full and compact errors The nature of XML documents complicates this analysis since violations can occur both in the XML document structure and in the XML Schema that governs policy content. When errors occur in the XML document structure, errors occurring later in the document are possibly masked since the syntactic interpretation of the document may change. If these analysis issues can be overcome, it would be valuable to determine if the identified errors were the result of poor policy development tools or issues with the P3P schema. With this information, it may be possible to improve the protocol either through tool development or future versions of the protocol itself. Further, if technologies such as P3P, which rely upon users to make reasonable choices regarding their privacy are to succeed, researchers must address the findings of Acusiti and Grossklags [31] and Spiekermann et al. [32] who found that users often trade long term privacy for short term benefits. If users can be persuaded to make reasonable choices, then technologies such as P3P may prove valuable. Another improvement which may be required is the development of user agents that continue to function even if P3P policies are not widely available. This might be accomplished through the collection of other observable server characteristics/documents or the analysis of third party information such as blacklists and reputation mechanisms.

It would also be of value to determine the extent that websites adhere to the privacy principles adopted by their respective jurisdictions. Such an analysis could conceivably be accomplished through the application of tools such as the JRC EU legislation analysis tool [40]. This tool could also be configured to allow for testing of other groups of websites such as those belonging to the US safe harbor program, or those belonging to other nations which have adopted privacy legislation such as Canada and Japan. Such an analysis may prove to be highly valuable since a growing body of evidence indicates that adherence to these laws/programs is questionable at best [66-68]. If such analysis is feasible, then P3P, or a derivative of P3P, may provide significant value as a tool for governments and organizations to remotely determine adherence to guidelines and legislation.

The analysis of P3P policies posted by organizations dealing with particularly sensitive information may also prove valuable. If members of the financial, health care, and data mining industries could be reliably identified, P3P could be used as a mechanism for remote analysis of their data handling practices thereby aiding public policy decisions. Finally, our results also suggest that the adoption of P3P may vary across cultural lines. If supported, this finding suggests that the design and development of P3P must begin to take into account cultural perceptions of privacy if they are to be implemented across the Internet.

2.9 Bibliography

[1] H. Kargupta, A. Joshi, K. Sivakumar, and Y. Yesha, *Data Mining: Next Generation Challenges and Future Directions*. Menlo Park, California: MIT Press, 2004.

[2] B. Malin, "Betrayed By My Shadow: Learning Data Identity via Trial Matching," *Journal of Privacy Technology,* 2004.

[3] H. Nissenbaum, "Protecting Privacy in an Information Age: The Problems of Privacy in Public," *Law and Philosophy,* vol. 17, pp. 559-596, 1998.

[4] J. Turow, L. Feldman, and K. Meltzer, "Open To Exploitation: American Shoppers Online and Offline," University of Pennsylvania's Annenberg School for Communication 2005.

[5] A. I. Antón, Q. He, and D. L. Baumer, "Inside JetBlue's Privacy Policy Violations," *IEEE security and privacy,* vol. 2, 2004.

[6] J. Rich, "Internet Service Provider Settles FTC Privacy Charges." vol. 2006: Federal Trade Commission, 2005.

[7] "On the Data Trail: How Detailed Information About You Gets Into The Hands of Organizations With Whom You Have No Relationship," The Canadian Internet Policy and Public Interest Clinic, Ottawa, Ontario, Canada April 2006.

[8] H. Black, "On-Line Data Brokers." vol. 2006: Office of the Privacy Commissioner of Canada, 2005.

[9] J. Rich, "ChoicePoint Settles Data Security Breach Charges; To Pay $10 Million in Civil Penalties, $5 Million for Consumer Redress." vol. 2006: Federal Trade Commission, 2006.

[10] "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data." vol. 2006, 1980.

[11] M. K. Landesberg, T. M. Levin, C. G. Curtin, and O. Lev, "Privacy Online: A Report to Congress." vol. 2006: Federal Trade Commission, 1998.

[12] "Safe Harbor Privacy Principles." vol. 2006: U.S. Department of Commerce, 2000.

[13] "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data," in *23.11.1995*, 1995.

[14] "The Platform for Privacy Preferences (P3P) Project." vol. 2006, 2002.

[15] P. Resnick and J. Miller, "PICS: Internet Access Controls Without Censorship," *Communications of the ACM,* vol. 39, pp. 87-93, 1996.

[16] C. Jensen and C. Potts, "Privacy Policies as Decision-Making Tools: An Evaluation of Online Privacy Notices," in *CHI 2004*, Vienna, Austria, 2004.

[17] OASIS, "OASIS eXtensible Access Control Markup Language (XACML) TC." vol. 2006, 2006.

[18] M. Schunter and C. Powers, "The Enterprise Privacy Authorization Language (EPAL 1.1)." vol. 2006: IBM, 2003.

[19] H. Hochheiser, "The Platform for Privacy Preferences as a Social Protocol: An Examination within the U.S. Policy Context," *ACM Transactions on Internet Technology,* vol. 2, pp. 276-306, November 2002.

[20] "Personal Information Protection and Electronic Documents Act," Second ed, 2000.

[21] "Act on the Protection of Personal Information," 2003.

[22] M. K. Reiter and A. D. Rubin, "Anonymous Web Transactions with Crowds," *Communications of the ACM,* vol. 42, pp. 32-48, February 1999.

[23] A. I. Antón, J. B. Earp, Q. He, W. Stufflebeam, D. Bolchini, and C. Jensen, "Financial Privacy Policies and the Need for Standardization," *IEEE security and privacy,* vol. 2, pp. 36-45, March-April 2004.

[24] C. Jensen, C. Potts, and C. Jensen, "Privacy Practices of Internet Users: Self-Reports Versus Observed Behavior," *International Journal Human Computer Studies,* vol. 63, pp. 203-227, 2005.

[25] L. F. Cranor, P. Gunuru, and M. Arjula, "User Interfaces for Privacy Agents," *ACM Transactions on Computer-Human Interaction,* 2006.

[26] T. Ledacky, H. Dhurvasula, and J. Loeb, "The Platform for Privacy Preferences (P3P)." vol. 2006: Mozilla.

[27] "Pretty Poor Privacy: An Assessment of P3P and Internet Privacy." vol. 2006: Electronic Privacy Information Center, 2000.

[28] K. Lee and G. Speyer, "Platform for Privacy Preferences Project (P3P) & Citibank," Citibank Advanced Development Group 1998.

[29] R. Thibadeau, "A Critique of P3P: Privacy on the Web." vol. 2006, 2000.

[30] "Privacy on the Internet - An Integrated EU Approach to On-line Data Protection," Article 29 - Data Protection Working Party, European Commission 5063/00/EN/FINAL WP 37, November 21 2000.

[31] A. AcQuisti and J. Grossklags, "Privacy and Rationality in Individual Decision Making," *IEEE security and privacy,* vol. 3, pp. 26-33, Jan-Feb 2005.

[32] S. Spiekermann, J. Grossklags, and B. Berendt, "E-Privacy in 2nd Generation E-Commerce: Privacy Preferences Versus Actual Behavior," in *3rd ACM conference on Electronic Commerce*, 2001.

[33] T. Yu, N. Li, and A. I. Antón, "A Formal Semantics for P3P," in *2004 Workshop on Secure Web Service*, Fairfax, Virgina, 2004, pp. 1-8.

[34] J. Linn, "Technology and Web User Data Privacy - A Survey of Risks and Countermeasures," *IEEE Security and Privacy,* vol. 3, pp. 52-58, 2005.

[35] L. Cranor, B. Dobbs, S. Egelman, G. Hogben, M. Langheinrich, M. Marchiori, M. Presler-Marshall, J. Reagle, M. Schunter, D. A. Stamply, and R. Wenning, "The Platform for Privacy Preferences 1.1 (P3P 1.1) Specification." vol. 2006: W3C, 2006.

[36] L. F. Cranor, S. Byers, and D. Kormann, "An Analysis of P3P Deployment on Commercial, Government and Children's Web Sites as of May 2003," Federal Trade Commission, Trade Report May 2003 2003.

[37] S. Byers, L. F. Cranor, and D. Kormann, "Automated Analysis of P3P-Enabled Web Sites," in *Fifth International Conference on Electronic Commerce*, Pittsburgh, 2003.

[38] L. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marshall, and J. Reagle,

"The Platform for Privacy Preferences 1.0 (P3P 1.0) Specification." vol. 2006: W3C, 2002.

[39] "Privacy Bird." vol. 2006.

[40] "JRC P3P Resource Center." vol. 2006: Joint Research Center.

[41] F. Arshad and S. Sheng, "Privacy Fox." vol. 2006, 2006.

[42] L. Cranor, M. Langheinrich, and M. Marchiori, "A P3P Preference Exchange Language 1.0 (APPEL 1.0)." vol. 2006: W3C, 2002.

[43] D. Julta and Y. Zhang, "Maturing e-Privacy with P3P and Context Agents," in *The 2005 IEEE International Conference on e-Technology, e-Commerce, and e-Services*, Hong Kong, 2005.

[44] "Alexa Traffic Rankings." vol. 2006, 2006.

[45] "FirstGov." vol. 2006.

[46] "Internet Retailer Top 400." vol. 2006, 2005.

[47] "Top 300 Guide." vol. 2006: Internet Retailer, 2004.

[48] "BBBonline." vol. 2006: Better Business Bureau, 2006.

[49] "Truste." vol. 2006: Truste.org, 2006.

[50] "Forbes 2000 Largest Public Companies." vol. 2006.

[51] "P3P Validator." vol. 2006, 2002.

[52] J. T. Casagrande, M. C. Pike, and P. G. Smith, "An Improved Approximate Formula for Calculating Sample Sizes for Comparing Two Binomial Distributions," *Biometrics,* vol. 34, pp. 483-486, September 1978.

[53] P. I. Good, *Resampling Methods: A Practical Guide To Data Analysis*, 3rd ed. Boston: Birkhauser, 2006.

[54] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences (2nd Edition).* Hillsdale, NJ: Lawrence Earlbaum Associates, 1988.

[55] J. Miller, "Statistical Significance Testing: A Panacea for Software Technology Experiments," *Journal of Systems and Software,* vol. 73, pp. 183-192, 2004.

[56] M. W. Lipsey and D. B. Wilson, *Practical Meta-Analysis* vol. 49. Thousand Oaks, California: Sage Publications, 2001.

[57] V. Hasselblad and L. V. Hedges, "Meta-Analysis of Screening and Diagnostic Tests," *Psychological Bulletin,* vol. 117, pp. 167-178, January 1995.

[58] "US E-Government Act of 2002, Public Law 107-347-DEC. 17 2002," 2002.

[59] J. Copas and D. Jackson, "A Bound for Publication Bias Based on the Fraction of Unpublished Studies," *Biometrics,* vol. 60, pp. 146-153, March 2004.

[60] T. Bay, J. Paoli, C. M. Sperberg-McQueen, E. Maler, and F. Yergeau, "Extensible Markup Language (XML) 1.0 (Third Edition)," 2004.

[61] J. Tian, S. Rudraraju, and L. Zhao, "Evaluating Web software reliability based on workload and failure data extracted from server logs," *IEEE Transactions on Software Engineering,* vol. 30, pp. 754-769, 2004.

[62] T. Huynh and J. Miller, "Further Investigations into Evaluating Web Site Reliability," in *4th International Symposium on Empirical Software Engineering,* Noosa Heads, Australia, 2005, pp. 162-171.

[63] "IP2Location." vol. 2006, 2006.

[64] N. Zakaria, J. M. Stanton, and S. T. M. Sarkar-Barney, "Designing and Implementing Culturally-Sensitive IT Applications," *Information Technology and People,* vol. 16, 2003.

[65] P. Beatty, I. Reay, S. Dick, and J. Miller, "P3P Adoption on E-Commerce Websites: A Survey and Analysis," *To appear in IEEE Internet Computing,* 2006.

[66] N. E. Bowie and K. Jamal, "Privacy Rights on the Internet: Self-Regulation or Government Regulation," *Business Ethics Quarterly,* vol. 16, July 2006.

[67] "Compliance with Canadian Data Protection Laws: Are Retailers Measuring Up?." vol. 2006: The Canadian Internet Policy and Public Interest Clinic, 2006.

[68] M. Markel, "Safe Harbor and Privacy Protection: A Looming Issue for IT Professionals," *IEEE Transactions on Professional Communication,* vol. 49, 2006.

Privacy Policies Versus National Cultural and Legislation on the Internet

3 Chapter 3 Introduction

When two people communicate, a significant part of the substance of their conversation is implicit, carried in the context and shared experiences of both individuals. This "common ground" provides a rich layer of meaning – one that can be completely misinterpreted when this understanding is *not* shared. Anyone who has had significant interactions with a different culture will have their own trove of amusing, frustrating, or tragic misunderstandings, brought about by the differing assumptions each side makes about the other's meaning or intent. Assumptions – and the expectations arising from them – are key elements of the human experience. They allow people to infer intentions, make decisions, and take action based on incomplete and imperfect information; they can be viewed as an individual's beliefs regarding the future outcome of a decision. For instance, a North American, who assumes that a store is competent, will have service quality expectations such as the ability to return goods found to be defective. In Hong Kong, however, stores usually treat sales as final, and so these individuals would not expect to be able to return defective goods, triggering changes in their behavior prior to purchasing an item. Such behavioral expectations form a mental model of how the brick-and-mortar store behaves.

In the online world, consumers are confronted with the same problems of incomplete and imperfect information, compounded by the risk of surrendering personal information to an entity (or entities) that the consumer has no personal interaction with. Consumers are therefore confronted with the need to determine if the privacy risks of a transaction outweigh the benefits of the transaction. There is a growing body of work showing that online consumers possess deeply flawed behavioral expectations [1-3] regarding website actions. For example, Turow *et al.* [1] found that:

- 59% of respondents believed incorrectly that "When a website has a privacy policy, it means that the site will not share my information with other websites or companies"
- 55% of respondents believed incorrectly that "When I give personal information to a bank, privacy laws say that the bank has no right to share that information, even with companies owned by the bank"
- 47% of respondents believed incorrectly that "When I give money to a charity, by law that charity cannot sell my name to another charity unless I give it permission"

We can interpret the findings in [1] as inaccurate mental models representing the actions of the online store. The lack of common ground between the online stores and the consumer, defined as the information or tools assumed to exist when they transact [4], is the source of these errors. Online stores commonly post privacy policies, which often explicitly state that they will undertake some or all of the actions studied in [1]; consumers, however, do not incorporate these facts into their mental models. Common ground concerning privacy in online transactions could include shared expectations of legal protections or business practices. When this common ground does not exist, it can be created through the standardization of business practices via standards development [5], creating laws [6, 7] or applying social pressure [8]. The current lack of common ground on privacy-sensitive actions damages consumer trust in online stores, which in turn impacts their willing to transact with online stores [9, 10].

Recent calls for standardization of privacy policies [5] and alignment of website privacy sensitive actions [2] attempt to address this issue. We interpret these calls as an effort to create norms of behavior, which consumers can then use as the basis of their expectations. The Internet, however, provides an environment where geographic, cultural, and legal restraints are marginalized, thus removing incentives to align and standardize practice. Findings by Shaw [11] and Hsu and Kuo [12] indicate that a wide range of cultural, legal, and organizational influences affect the privacy sensitive decisions made by employees of organizations, while consumer's privacy preferences are also culturally dependent [13, 14]. These findings suggest that cultural differences may significantly impact the behavioral norms that website operators can or will adhere to; not all options will be palatable to organizations in different nations, or to the consumers they wish to attract! To date, however, there is little empirical evidence on the impact of cultural differences on the privacy sensitive actions of website operators.

The goal of our research is to analyze the privacy-sensitive decisions made by operators of the Internet's most popular websites. Alexa.com has provided us with a copy (circa November 2005) of their list of the 100,000 most popular sites on the Internet [15]. Through this analysis, we will provide broad-based empirical evidence on: cultural differences in the adoption of privacy enhancing technologies, cultural differences in privacy-sensitive actions, the degree of standardization of stated practice within a culture, and the effectiveness of legal frameworks in creating norms of behavior within a culture. This evidence will be important in the future development of privacy enhancing technologies, legal frameworks, and public education concerning online privacy protection. Due to the magnitude of this survey, we use an automated approach, based on the Platform for Privacy Preferences (P3P) [16] documents posted on surveyed websites. P3P is a significant privacy-enhancing technology, which provides a machine-readable version of a website's privacy policy – making these documents far more amenable to analysis than natural-language privacy policies.

The remainder of this paper is organized as follows: In Section 3.1, we overview the current state of online privacy. In Section 3.2, we discuss the relationship between culture and privacy. In Section 3.3, factors influencing the privacy sensitive decisions of organizations are discussed, and we frame our hypotheses concerning the interaction of culture and legislation with these actions. In Section 3.4, we describe our survey and analytical methodology. We describe the results of our analysis in Section 3.5. In Section 3.6, an exploratory analysis across one particular cultural dimension reveals some important differences between cultures. In Section 3.7, we study the structure of privacy-sensitive actions for each culture, searching for clusters of website behaviors that consumers could potentially use in the absence of overall norms. Finally, we provide a summary and discussion of future work in Section 3.8.

3.1 Current State of Online Privacy

To date, a wide range of social and technological methods have been proposed in an attempt to protect a website patron's privacy. Currently, these methods include the use of human readable privacy policies (HRPP) [5, 17, 18], third party assurance services (TPA) [19-22], and privacy enhancing technologies (PETs) such as P3P [16], anonymization tools [23, 24], cookie managers [23, 25], and encryption [24, 26]. The social mechanisms (HRPPs and TPA) provide information to a consumer, enabling them to form accurate behavioral expectations about the practices of a specific website. On the other hand, PETs (anonymization, cookie managers and encryption) require the user to develop a *desired* mental model, and then assist the user in *enforcing* that model in their online transactions. The common ground created in this latter case is essentially adversarial rather than cooperative, but it does result in fairly accurate behavioral expectations. While these mechanisms can develop common ground between the consumer and website operator, recent empirical evidence (discussed below) indicates that these mechanisms are insufficient for the *average* user to develop accurate behavioral expectations.

HRPPs are intended to inform website patrons of the privacy sensitive actions undertaken by a website. They are constructed using legal terminology, as a posted privacy policy is legally enforceable in many jurisdictions. The terminology that is used is thus opaque to most users [17]. Additionally, HRPP's also lack a standardized format, are rarely read [27, 28], time consuming to analyze [2], and potentially deceptive [18]. These findings suggest that HRPP's currently do not generate common ground between vendors and consumers, and do not solve the problem of privacy protection.

TPA services are another attempt to solve many of the aforementioned usability problems with HRPP's. These services create a standard to which member organizations adhere. Often TPA services provide a trademarked certification to organizations, which can be posted on their websites proving they satisfy the provider's requirements. Truste [20], BBBOnline [19], CPA Web Trust [29], and Verisign [30] represent the most popular TPA authorities. While TPA services provide a means of describing what actions are undertaken by organizations, they require individuals to be aware of the certification's existence, implications, and limitations. Since only a minority of individuals can identify the Truste (42% of individuals), BBBOnline (28.7%), or CPA Web Trust (7.7%)

certifications [31], it would appear unlikely that Internet users can discern the differences between certifications let alone interpret their meaning.

In contrast to the predominantly social mechanisms that have been described, a number of privacy-enhancing technologies (PET) have been proposed, including anonymization services, cookie management, and encryption tools (P3P is often referred to as a PET; we discuss this further below). However, usability issues plague their widespread adoption since these mechanisms currently demand a high degree of technical sophistication from the user. PETs are rarely part of a standard system configuration, require some knowledge of the threat model behind the technology, and frequently demand significant effort to configure. For example the MozPETs browser plug-in [23] for the Firefox web browser requires individuals to:

1) Know that cookies exist
2) Determine the vulnerabilities associated with different classes of cookies
3) Create detection rules (possibly using regular expressions! [32])

Given that the average Internet user cannot explain the basic functionality of cookies [28], these requirements are unrealistic for the general population.

P3P [16] is an attempt to provide a machine readable form of a websites HRPP. In the P3P protocol, HRPPs are mapped to a machine readable format by converting the semantic contents of a privacy policy into an XML document by means of an XML schema provided in the P3P specification [33]. Once a HRPP is translated into a P3P document and posted on a website, these documents are retrieved and analyzed by software (within the user's browser) in the hope of providing the user with a salient summary of the policy.

P3P agents are also able to warn users when a privacy policy does not conform to the user's preferences. For this reason, P3P can be viewed as both a social and a technological mechanism; it both provides information for, and supports a user in, enforcing behavioral expectations. While the P3P protocol appears to be an improvement over HRPPs, it has not been widely adopted by websites. Researchers have found general P3P adoption to be approximately 9% and 10% in May and June of 2003 respectively [34, 35]. Recent follow up studies [36, 37] describe little if any change in current adoption rates (with the exception of US government websites, likely due to the recently enacted E-Governance Act [38]). This adoption rate, while sufficient for a survey analyzing website actions such as ours, limits P3P's impact as a privacy enhancing technology.

Since the methods above appear incapable of creating sufficient common ground between website operators and average users for accurate behavioral expectations to develop, a number of nations have enacted privacy-protection legislation [39, 40]. While these laws have the potential to standardize website practices, it is only reasonable to expect this standardization to exist within a legal jurisdiction. Since international consensus on the most suitable legal framework does not yet exist [39], we cannot currently expect privacy protection to be standardized Internet- (or world-) wide. In addition, further complications arise when legal frameworks that are similar in spirit differ in written content and enforcement measures [39, 41]. As a result of these framework inconsistencies, a range of protections exists. For example, while the United States has enacted privacy laws, these laws are restricted to a few select sectors of industry [42]. In contrast, nations belonging to the European Union [43], Canada [44], and Japan [45] have enacted laws which govern both public and private organizations (these laws will be referred to as 'general privacy laws'). These differences tend to imply that only some jurisdictions would possess legal protections sufficiently stringent as to foster the development of standardized practice.

This legal environment is further complicated by the findings of Turow *et al.* [1] which indicate that even where laws exist, individuals are unaware of the protections afforded. This suggests that even where legal protections are provided, individual users are unlikely to be able to utilize them. These conjectures imply that standardized practice is likely to exist only where an effective third party enforces the legal framework. To date, third party enforcement is implemented through the creation of Ombudsmen [46], registration offices [47], or licensing bureaus [48].

3.2 Culture and Privacy

Cultural differences have been shown to affect individuals' expectations [49], privacy concerns [50, 51], social norms, and the laws of a nation [6]. These differences are often analyzed through comparisons of national populations whereby national boundaries are assumed to be synonymous with cultural boundaries [14, 49, 52, 53]. This assumption is reasonable for this survey since privacy legislation is predominantly enacted at a national level. However, large multi-ethnic nations such as the U.S. or China are a reminder of the limitations of this assumption; there may be many cultures within such a nation, which may hold significantly different beliefs, values and norms than the majority. Hofstede's cultural dimensions [54] are frequently utilized to describe the differences between national cultures. Through extensive empirical research, Hofstede has identified five cultural dimensions:

- *Individualism*: focuses on the degree the society reinforces individual or collective achievement and interpersonal relationships

- *Power Distance*: focuses on the degree of equality, or inequality, between people in the country's society.
- *Masculinity*: focuses on the degree the society reinforces, or does not reinforce, the traditional masculine work role model of male achievement, control, and power.
- *Uncertainty Avoidance*: focuses on the level of tolerance for uncertainty and ambiguity within the society - i.e. unstructured situations.
- *Long Term Orientation*: focuses on the degree the society embraces, or does not embrace, traditional values, deferred gratification, and long-term commitments.

In addition, Milberg *et al.* [50] have found associations between Hofstede's *Individualism*, *Power Distance*, *Masculinity*, and *Uncertainty Avoidance* dimensions and a culture's concern for information privacy, justifying the analysis of privacy topics using culturally sensitive approaches.

While Hofstede's cultural dimensions explain how cultures differ conceptually, they provide little insight into how individuals communicate and form expectations. Hall's theories of High and Low-Context communication [55] provide insight into this issue. Hall hypothesizes that communication is context dependent and that this context varies between cultures. Individuals from 'High-Context' cultures utilize extensive implicit information (common ground) to communicate knowledge, expecting others to utilize this context in interpreting the communication. In contrast, individuals from 'Low-Context' cultures assume less common ground between individuals, leading to communication that is more verbose. These results tend to imply that individuals from cultures may prefer to explain their actions through differing mediums. For example, individuals from Low-Context cultures are more accustomed to verbose documents (such as human readable privacy policies) than their High-Context counterparts.

The cultural dimensions identified by Hall and Hofstede have both been widely utilized as well as strongly criticized [56] as a result of their rough categorizations of 'national culture'. Ess and Sudweeks [57] go so far as to state "having only five or six dimensions for the analysis of culture seems like attempting brain surgery with a bulldozer"; yet they also conclude that in many instances, the dimensions of Hall and Hofstede possess sufficient predictive and explanatory power. In summary, the use of Hall and Hofstede's indices is considered appropriate so long as models or inferences based on them are statistically significant.

3.3 Culture and Privacy Sensitive Actions Within Organizations

Shaw [11] and Hsu and Kuo [12] have shown that the Theories of Moral Intensity and Planned Behavior can explain the privacy sensitive actions taken by individuals within organizations. In particular, they found evidence that social norms and the perceived magnitude of the resulting effects influence individuals' decisions. Both of these concepts are culturally dependent. Cultural factors predispose an individual to adhere to certain social norms. Culture also affects an individual's perception of the risks related to an action, thus modifying the magnitude of the action's resulting effect. For instance, some advertisers in the United States (low *uncertainty avoidance*) currently utilize 'spyware' for the delivery of directed advertising [58]. While certainly unethical and possibly illegal, these actions have *so far* not resulted in penalties levied against the company in question. Individuals from cultures with a *high uncertainty avoidance* score are unlikely to accept the inherent risk of crippling civil or even criminal penalties in this strategy.

Based upon the theories outlined in the previous sections, we propose several hypotheses concerning cultural influences on P3P adoption and policy content. First, given that individuals, and by extrapolation organizations, are influenced by culture, it would appear reasonable to hypothesize that the actions of Internet organizations will vary between cultures. Further, since individuals from Low-Context cultures may be more inclined to the usage of legal documents, it would appear reasonable to expect the adoption of HRPPs, including specifically P3P policies, to be higher in Low-Context cultures.

> ***Hypothesis 1:*** *A statistically significant increase in P3P adoption rates will exist in websites from Low-Context nations when compared to websites from High-Context nations.*

It would also appear reasonable to hypothesize that an increase in a cultures' concern for information privacy should be associated with their adoption of privacy protection technologies, including P3P. Thus, we propose the following hypotheses based upon the findings of Milberg *et al.* [50] (which excluded Hofstede's *Long-Term Orientation* dimension):

> ***Hypothesis 2:*** *A statistically significant positive association should be identified between P3P adoption and a culture's Individualism Dimension.*

> ***Hypothesis 3:*** *A statistically significant positive association should be identified between P3P adoption and a culture's Masculinity Dimension.*

> ***Hypothesis 4:*** *A statistically significant positive association should be identified between P3P adoption and a culture's Power Distance Dimension.*

> ***Hypothesis 5:*** *A statistically significant negative association should be identified between P3P adoption and a culture's Uncertainty Avoidance Dimension.*

If supported, Hypotheses 1-5 provide empirical evidence on the relationship between cultural dimensions and the adoption of PETs, and guidance on which cultural factors are most important in the decision to adopt a PET. This would be a significant finding for the PET community, in that careful attention to cultural sensitivity would be required in designing new PETs. In addition to this analysis, we also seek evidence on the emergence or non-emergence of standards of behavior, and whether the laws of a nation influence the emergence of such standards. This evidence would be valuable in the development of future legal frameworks and in public education efforts, in addition to revealing the impact of cultural differences on the privacy sensitive actions of website operators.

Since the actions of employees (and their organizations) are partially governed by their perceptions of the magnitude of privacy effects, it would appear unreasonable to expect companies to undertake similar practices without external influences such as legislation or strong public opinion. However, as previously indicated, Internet users in general do not appear to have an adequate grasp of Internet privacy issues to mobilize a “grassroots” social campaign for the creation of standard practices. Thus, legislative action appears to be the only realistic means of influencing organizational choices. This influence is however contingent upon legislation being sufficiently stringent, as well as the existence of a legal framework providing effective third party enforcement.

While legislation has the potential to establish standard practice, it is unrealistic to expect complete uniformity in the stated actions of websites since transaction context varies. For example, it is reasonable for an E-Commerce company to request your address, retain it until shipping is completed, and share the address with a shipping company. Similar actions undertaken by a search engine are likely to be considered unwarranted. As a consequence, associations should only be expected for certain sub-categories of P3P content that are often described in legislation [39, 43-45], such as accessibility of collected information, dispute resolution methods employed, remedies offered when problems occur, and *transaction independent* purposes for data collection such as administering the website, profiling individuals, telemarketing, etc. For these reasons, we propose the following hypotheses:

Hypothesis 6a: *For each nation utilizing general privacy laws with third party enforcement, a significant association with respect to whether individuals can access information collected about them should exist.*

Hypothesis 6b: *For each nation not utilizing general privacy laws with third party enforcement, a significant association should not exist as to whether individuals can access information collected about them.*

Hypothesis 7a: *For each nation utilizing general privacy laws with third party enforcement, a significant association regarding dispute resolution methods offered should exist.*

Hypothesis 7b: *For each nation not utilizing general privacy laws with third party enforcement, a significant association should not exist regarding dispute resolution methods offered.*

Hypothesis 8a: *For each nation utilizing general privacy laws with third party enforcement, significant association as to the remedies offered should exist.*

Hypothesis 8b: *For each nation not utilizing general privacy laws with third party enforcement, a significant association should not exist as to the remedies offered.*

Hypothesis 9a: *For each nation utilizing general privacy laws with third party enforcement, a significant association regarding transaction independent purposes for collection should exist.*

Hypothesis 9b: *For each not nation utilizing general privacy laws with third party enforcement, a significant association should not exist regarding transaction independent purposes for collection.*

Hypothesis 10a: *For each nation, a significant association should not exist regarding categories of data collected.*

Hypothesis 10b: *For each nation, a significant association should not exist regarding how long information is retained.*

Hypothesis 10c: *For each nation, a significant association should not exist regarding who is permitted access to information.*

3.4 Survey Methodology

The survey was completed in November of 2005 and consisted of an automated analysis of P3P documents posted on the most popular 100,000 Internet websites as ranked by Alexa.com [15]. P3P policies were chosen as the unit of analysis due to the extreme time commitments required to analyze human readable privacy policies. An analysis of 100,000 human readable privacy policies using a method analogous to Earp *et al.* [2] would take approximately 800 person years to complete and is clearly not cost-effective.

3.4.1 Website Selection

The choice of using P3P policies from the Alexa 100,000 list [15] instead of a random sample originates from a desire to analyze business practices from a user perspective and is in accordance with the survey methodology of previous P3P surveys [34, 35, 37, 59].

The websites contained in this list comprise the vast majority of websites a user is likely to encounter and consequently have the greatest affect upon an average individual's privacy. Websites excluded from this list have less than a 0.00125% chance that the average Internet user will visit them [15]. Further, the use of the Alexa 100,000 list allows for an analysis that draws upon websites from 131 countries. Hence, we view this list as a highly representative sample of the "usable" Internet. In fact, it could be argued that the list effectively represents the entire "usable" Internet.

3.4.2 P3P Policy Harvesting

The P3P validator [60] is the official W3C P3P document verification tool. It is openly provided and provides a convenient method for retrieving, parsing, and validating retrieved P3P documents. The P3P validator locates policies using the official methods defined in the P3P specification [33] such as the use of HTTP headers, HTML link tags, and the well-known location (/w3c/p3p.xml). Once the documents are retrieved and parsed, Perl scripts export the results to a MySQL database [61] where information is stored as binary values (whether the tag exists or not; categorical XML tag attributes are dichotomized). A reverse DNS look-up was performed for each website using the Linux 'host' program. Once the IP address was retrieved, a database purchased from IP2Location [62] was used to locate the corresponding country of origin. IP2Location states their accuracy is above 95% [62].

3.4.3 Data Analysis

To test hypothesis 1, a one-way equal weighted Chi-Square test [63] analyzing the dichotomous existence of P3P policies posted on websites from High versus Low-Context nations is utilized. The equal weighted Chi-Square test was employed to ensure that sample size differences in the various populations of websites did not bias the results.

For all significance tests, α is set to the traditional 0.05 level. In order to determine the operational effect of any observed statistically significant differences, we also employ effect size tests. All effect size tests will be reported as standardized mean difference effect sizes, commonly known as Cohen's d [64, 65]. The standard method for calculating Cohen's d effect sizes from dichotomous contingency tables involves either probit or arcsine transformation proportions [65]. However, these methods assume that the observed dichotomous variables represent normally distributed populations, thus making them inappropriate for testing for the existence of a trait [64, 65]. As a result, the operational effect pertaining to hypothesis 1 will be calculated by an application of an odds ratio effect size test [65] whose result can be converted into a Cohen's d effect size using the method outlined by Hasselblad and Hedges [66]. The magnitude of the operational effect will be determined by following the guidance provided by Cohen where $d \geq 0.2$ indicates a small effect, $d \geq 0.5$ is a medium effect, and $d \geq 0.8$ indicates a large effect [64, 65].

Hypotheses 2-5 will be tested through an application of Spearman's Rank Correlation test [67]. The observed P3P adoption frequency (represented as a percentage of the total national websites served from a nation) will be compared against the nation's score for a particular cultural dimension as identified by Hofstede [54]. A non-parametric test is required due to the abstraction of dichotomous P3P adoption variables into a frequency form. The operational effect will be determined by calculating the Cohen's d effect size from the correlation coefficient [65]. With respect to hypotheses 6a-10c, our goal is to determine the degree of association within websites from an individual jurisdiction. Due to the potentially large number of variables, standard bivariate approaches such as

Pearson or Spearman Rank correlations are not applicable. For these reasons, Cronbach's Alpha [68, 69] and Intraclass Correlation [67, 70-73] methods are utilized. Cronbach's Alpha provides a method for rating the internal consistency of a series of responses (website P3P policies) on various subjects of interest (P3P tags). However, Cronbach's Alpha is not a measure of unidimensionality. A dataset can have a high Cronbach's Alpha and still be multidimensional when clusters of highly intercorrelated items exist and weak correlation exists between these clusters [74]. As a result, Cronbach's Alpha will be utilized as an exploratory test indicating whether a population exhibits general internal consistency. Thus, while a weak result indicates no consistency or agreement, a strong result cannot be interpreted as agreement between websites regarding privacy issues. The lower bound of 0.7, stated by Nunnally and Bernstien [75], will be used as an indication of general internal consistency.

Intraclass correlation will be utilized to determine the degree of absolute website agreement on the subjects of interest (P3P tags). Intraclass correlation (ICC) should not be confused with standard correlation techniques; ICC allows for analysis concerning both the degree of association as well as the repeatability of the association when the observed variables are dichotomous in nature. The analysis of this paper utilizes the method generally referred to as Model 3 [71, 73] using individual measurements; Model 3 is appropriate when all subjects of interest are surveyed [71]. This approach is reasonable since this study surveys all popular P3P adopting websites. Generalizations to all popular websites (whether they adopt P3P or not) would require the usage of Model 2 and must be made at the reader's discretion. However, as a cross check, the ICC results for Model 2 were calculated and no significant differences exist between the results from Models 2 and 3.

Interpretations of ICC results differ from those generated through standard correlation methods; the ICC value cannot be viewed as an r^2 value since it takes into account consistency *as well as* repeatability of results. We will follow the recommendations of Portney and Watkins [Portney and Watkins, 1993], who state that an ICC above 0.75 is an indicator of good agreement. In practice, ICC values range from 0.00 to 1.00. However, in certain circumstances, the ICC values can range from ±∞ [Portney and Watkins, 1993]. This situation occurs when the data set is homogeneous, meaning a lack of significant variance between subjects of interest (P3P tags); this situation can be discovered by the application of a one-way ANOVA test [67]. When a dataset is found to be homogenous, the test is considered inconclusive.

A limitation of using the ICC method is the lack of clear consensus regarding the calculation of Type I and Type II errors. For example, little guidance exists regarding the choice of a null hypothesis for an ICC test. The choice of $H_0 = 0$, for instance, provides information of little practical importance. Walter *et al.* [76] recommended setting H_0 to the minimally acceptable level of reliability. However, little guidance is given for choosing this point without making the tests overly conservative. The calculation of a minimum detectable effect (MDE) [77] is a potential solution. MDE's describe the smallest effect that can be detected at a given statistical power and significance level, thus guarding against these errors. However, no formulation exists for the calculation of a MDE when using ICC. As a result of these limitations, confidence intervals will be calculated as an indicator of the degree of uncertainty in the results [78]. Further, this limits our ability to provide insight into the hypotheses 6b,7b,8b,9b,10a,10b and 10c; and hence, we will only make very conservative statements about our findings with regard to these hypotheses.

3.5 Survey Analysis

Of the 100,000 websites provided by Alexa.com [15], 97,418 were reachable at the time of the survey. This response rate of 97.4% is similar to the 97.8% [35] and 98.0% [34] response rates observed in previous studies analyzing P3P adoption. Of the 97,418 reachable websites, 2,345 posted error free P3P policies, or 2.4% of the survey.

3.5.1 Analysis of Hypotheses 1-5

Due to the restricted range of nations analyzed by Hall [55, 79] and Hofstede [54, 79], the presented analysis cannot encompass all 131 nations in the Alexa list. The subset of nations which appear in Hofstede's surveys, Hall's classifications and possess at least a single surveyed website, constitutes 52 nations and is depicted in Table 3-1.

Through analysis of Table 3-1, a statistically significant (weighted chi-square 27.576, $p <$ 0.001) difference in P3P adoption with a large (d = 1.294) effect size was identified between High and Low-Context nations, supporting Hypothesis 1. This finding is important, since it empirically supports the findings of Shaw [11] and Hsu and Kuo [12] with data analyzing stated organizational actions, rather than only personal perceptions of employees and supports the formulation of the other culturally sensitive hypotheses in this paper.

Table 3-1: The Largest Subset of Nations Consistent in Our Survey, Hofstede's Survey, and Hall's Classification. (Part 1)

Our Survey				Hofstede's Scores				Halls Classification
Country	Total Available	Total number of P3P enabled sites	Percentage of P3P enabled Sites	Power Distance	Individualism	Masculinity	Uncertainty Avoidance	High or Low Context
New Zealand	120	9	7.50%	22	79	58	49	Low
Costa Rica	43	3	6.98%	35	15	21	86	High
Belgium	238	14	5.88%	65	75	54	94	Low
Denmark	422	22	5.21%	18	74	16	23	Low
United Kingdom	3481	174	5.00%	35	35	89	66	low
Netherlands	1501	63	4.20%	38	80	14	53	Low
Canada	2532	90	3.55%	39	80	52	48	Low
United States	44591	1577	3.54%	40	91	62	46	Low
Germany	2411	80	3.32%	35	67	66	65	Low
Australia	769	25	3.25%	36	90	61	51	Low
South Africa	70	2	2.86%	60	65	39	85	Low
Chile	80	2	2.50%	63	23	28	86	High
Argentina	173	4	2.31%	49	46	56	86	High
France	1845	40	2.17%	68	71	43	86	Low
Sweden	693	15	2.16%	31	71	5	29	Low
Switzerland	284	6	2.11%	34	68	70	58	Low
Norway	258	5	1.94%	31	69	8	50	Low
Malaysia	160	3	1.88%	104	26	50	36	High
Finland	168	3	1.79%	33	63	26	67	Low
Brazil	401	6	1.50%	69	38	49	76	High
Italy	638	9	1.41%	50	76	70	75	Low
Israel	671	9	1.34%	13	54	47	81	Low
Portugal	99	1	1.01%	63	27	31	104	High
Greece	320	3	0.94%	60	35	57	112	High
Japan	6743	55	0.82%	54	46	95	92	High
South Korea	1724	14	0.81%	60	18	39	85	High
Turkey	738	5	0.68%	66	37	45	85	High
Taiwan	1301	5	0.38%	58	17	45	69	High
Thailand	312	1	0.32%	64	20	34	64	High
Hong Kong	1398	3	0.21%	68	25	57	29	High
China	17043	6	0.04%	80	20	66	30	High
Colombia	30	0	0.00%	67	13	64	80	High
Ecuador	10	0	0.00%	78	8	63	67	High
Egypt	111	0	0.00%	80	38	52	68	High
Guatemala	4	0	0.00%	95	6	37	101	High
India	223	0	0.00%	77	48	56	40	High

Table 3-1: The Largest Subset of Nations Consistent in Our Survey, Hofstede's Survey, and Hall's Classification. (Part 2)

Our Survey				Hofstede's Scores				Halls Classification
Country	Total Available	Total number of P3P enabled sites	Percentage of P3P enabled Sites	Power Distance	Individualism	Masculinity	Uncertainty Avoidance	High or Low Context
Indonesia	63	0	0.00%	78	14	46	48	High
Iran, Islamic Republic Of	13	0	0.00%	80	38	52	68	High
Mexico	176	0	0.00%	81	30	69	82	High
Pakistan	13	0	0.00%	55	14	50	70	High
Panama	31	0	0.00%	95	11	44	86	High
Peru	29	0	0.00%	64	16	42	87	High
Philippines	31	0	0.00%	94	32	64	44	High
Singapore	214	0	0.00%	74	20	48	8	High
Uruguay	75	0	0.00%	61	36	38	100	High
Venezuela	69	0	0.00%	81	12	73	76	High
Kuwait	20	0	0.00%	80	38	52	68	High
Saudi Arabia	128	0	0.00%	80	38	52	68	High
Syrian Arab Republic	17	0	0.00%	80	38	52	68	High
United Arab Emirates	65	0	0.00%	80	38	52	68	High
Nigeria	1	0	0.00%	77	20	46	54	High
Ireland	104	0	0.00%	28	70	68	35	Low

A statistically significant ($p < 0.001$) positive correlation with a large (d = 1.384) effect size was found for Hypothesis 2. This result corresponds with the finding of Milberg *et al. [50]*. Hypothesis 4 however, presents a different picture, where a statistically significant ($p < 0.001$) negative correlation with a large (d = 1.626) effect size, was identified between P3P adoption and *Power Distance* scores. Milberg *et al.* [50] identified a positive correlation with a medium (d = 0.59) effect size in their study. While these findings provide further support for the theories of Shaw [11] and Hsu and Kuo [12], they also suggest that a cultures concern for information privacy and it's general adoption of privacy enhancing technologies may not be strongly correlated. Further support for this conjecture comes from the lack of statistically significant correlations between P3P adoption and Hofstede's dimensions of *Masculinity* (hypothesis 3, $p <$ 0.221) and *Uncertainty Avoidance* (hypothesis 5, $p < 0.670$) whereas Milberg *et al.* [50] identified large (d = 1.22) positive and medium (d = 0.69) negative, statistically significant, correlations respectively. If supported, this finding indicates that there is a discrepancy between concern for information privacy in a culture (examined by Milberg *et al.* [50]), and the actual adoption of PETs in a culture. There is some apparent similarity between this discrepancy and the discrepancy between *individual's* stated and actual behaviors on the Web [80]; an exploration of this similarity is however beyond the scope of our present study.

3.5.2 Analysis of Hypotheses 6a-10c

In order to test hypotheses 6a-10c a reduced set of nations must be analyzed, since the vast majority of nations in Table 3-1 contain few if any websites utilizing the P3P protocol. In addition, nations not appearing in Hofstede and Hall's surveys that have P3P enabled websites are re-integrated back into the analysis. Accordingly, the analysis utilizing ICC methods is depicted in Tables 3-2 to 3-9. These 15 nations had the highest P3P adoption rate and constitute 95% of all the observed P3P policies. The analysis of hypotheses 6a-10c also requires the identification of whether general privacy laws and third party enforcement mechanisms exist in the various nations under study. General privacy laws with third party enforcement mechanisms have been implemented in the Netherlands [81, 82], Belgium [83, 84], Spain [85, 86], Sweden [87, 88], United Kingdom [47, 89], Denmark [90, 91], France [48, 92], Germany [93, 94], Australia [95, 96], Canada [44, 46], and New Zealand [97, 98]. Japan possesses general privacy laws

[45], but does not implement a third party protection agency. The USA and Korea both implement sector specific legislation [42, 99]. Russia only provides constitutional privacy protections [100].

The results of Table 3-2 indicate that as expected (with the exception of Russia), total agreement in P3P document content was not observed. It is very surprising that Russian websites, which are governed by very weak privacy protections [100, 101] showed significant agreement. This agreement could be a result of an unidentified relationship either through ownership or industry sector. Another plausible independent explanation could be that due to the lack of legal protections, Russian websites all report that they undertake any action they wish. This unique agreement in Russian websites will be analyzed further in the following sections.

Table 3-2: Intraclass Correlation Between All Privacy Related P3P Tags (53 Total Tags)

				95% confidence interval		F-Test		
Country	Number Of P3P policies	Cronbach's Alpha	Intraclass Correlation	lower	Upper	Value	df	*p*
Nations With General Privacy Laws and Third Party Protection								
Australia	25	0.933	0.329	0.248	0.438	14.8	52	< 0.001
Belgium	14	0.913	0.424	0.328	0.541	11.5	52	< 0.001
Canada	91	0.977	0.296	0.226	0.394	42.9	52	< 0.001
Denmark	22	0.897	0.249	0.178	0.349	9.7	52	< 0.001
France	40	0.970	0.436	0.348	0.545	33.3	52	< 0.001
Germany	80	0.978	0.341	0.265	0.444	45.2	52	< 0.001
Netherlands	63	0.974	0.346	0.268	0.451	37.8	52	< 0.001
New Zealand	9	0.916	0.526	0.419	0.642	11.9	52	< 0.001
Spain	15	0.929	0.453	0.357	0.569	14.0	52	< 0.001
Sweden	15	0.908	0.343	0.252	0.460	10.8	52	< 0.001
United Kingdom	176	0.989	0.317	0.245	0.417	89.7	52	< 0.001
Nations Without General Privacy Laws or Third Party Protection								
Japan	56	0.967	0.308	0.235	0.410	30.1	52	< 0.001
Korea	14	0.726	0.130	0.077	0.210	3.7	52	< 0.001
Russian Federation	24	0.994	0.863	0.813	0.907	174.9	52	< 0.001
USA	1586	0.996	0.298	0.229	0.394	239.3	52	< 0.001

Table 3-3: Intraclass Correlation Between All Access Related P3P Tags (6 total tags)

				95% confidence interval		F-Test		
Country	Number Of websites	Cronbach’s Alpha	Intraclass Correlation	lower	upper	Value	df	*p*
Nations With General Privacy Laws and Third Party Protection								
Australia	25	0.752	0.127	0.024	0.529	4.0	5	.002
Belgium	14	0.488	0.076	-0.026	0.484	2.0	5	.098
Canada	91	0.887	0.093	0.031	0.407	8.8	5	< 0.001
Denmark	22	0.884	0.293	0.107	0.737	8.6	5	< 0.001
France	40	0.891	0.196	0.069	0.620	9.1	5	< 0.001
Germany	80	0.960	0.266	0.115	0.694	25.2	5	< 0.001
Netherlands	63	0.908	0.158	0.057	0.552	10.9	5	< 0.001
New Zealand	9	0.911	0.578	0.278	0.902	11.3	5	< 0.001
Spain	15	0.821	0.269	0.0.76	0.726	5.6	5	< 0.001
Sweden	15	0.879	0.367	0.138	0.798	8.2	5	0.001
United Kingdom	176	0.915	0.068	0.024	0.322	11.7	5	< 0.001
Nations Without General Privacy Laws or Third Party Protection								
Japan	56	0.845	0.104	0.030	0.447	6.4	5	< 0.001
Korea	14	0.872	0.369	0.135	0.800	7.8	5	< 0.001
Russian Federation	24	0.995	0.915	0.800	0.985	217.0	5	< 0.001
USA	1586	0.979	0.101	0.041	0.409	49.0	5	< 0.001

Table 3-4: Intraclass Correlation Between All Dispute Resolution P3P Tags (4 total tags)

				95% confidence interval		F-Test		
Country	Number Of websites	Cronbach's Alpha	Intraclass Correlation	lower	Upper	Value	Df	*p*
Nations With General Privacy Laws and Third Party Protection								
Australia	25	0.987	0.795	0.534	0.982	77.9	3	< 0.001
Belgium	14	0.836	0.295	0.061	0.874	6.1	3	0.021
Canada	91	0.995	0.702	0.425	0.971	196.2	3	< 0.001
Denmark	22	0.934	0.416	0.154	0.914	15.2	3	< 0.001
France	40	0.992	0.782	0.522	0.981	118.2	3	< 0.001
Germany	80	0.996	0.770	0.512	0.979	250.0	3	< 0.001
Netherlands	63	0.991	0.673	0.388	0.967	112.6	3	< 0.001
New Zealand	9	0.853	0.449	0.108	0.930	6.8	3	0.002
Spain	15	0.982	0.786	0.511	0.981	56.0	3	< 0.001
Sweden	12	0.959	0.639	0.321	0.963	24.6	3	< 0.001
United Kingdom	176	0.998	0.768	0.513	0.979	535.5	3	< 0.001
Nations Without General Privacy Laws or Third Party Protection								
Japan	56	0.972	0.406	0.168	0.907	35.7	3	< 0.001
Korea	14	0.598	0.092	-0.018	0.697	2.5	3	0.075
Russian Federation	24	0.996	0.913	0.761	0.993	253.0	3	< 0.001
USA	1586	0.998	0.593	0.318	0.953	658.7	3	< 0.001

Table 3-5: Intraclass Correlation Between All Remedy Related P3P Tags (3 total tags)

				95% confidence interval		F-Test		
Country	Number Of websites	Cronbach's Alpha	Intraclass Correlation	Lower	upper	Value	df	*p*
Nations With General Privacy Laws and Third Party Protection								
Australia	25	0.990	0.787	0.480	0.993	91.2	2	< 0.001
Belgium	14	0.798	0.198	0.013	0.924	4.9	2	0.015
Canada	91	0.996	0.738	0.427	0.991	231.3	2	< 0.001
Denmark	22	0.935	0.342	0.096	0.956	15.5	2	< 0.001
France	40	0.993	0.804	0.513	0.994	151.6	2	< 0.001
Germany	80	0.989	0.537	0.231	0.976	87.1	2	< 0.001
Netherlands	63	0.967	0.356	0.117	0.958	30.6	2	< 0.001
New Zealand	9	0.865	0.517	0.089	0.980	7.4	2	< 0.001
Spain	15	0.928	0.420	0.120	0.969	13.8	2	< 0.001
Sweden	15	0.962	0.593	0.241	0.984	26.0	2	< 0.001
United Kingdom	176	0.998	0.718	0.405	0.990	412.2	2	< 0.001
Nations Without General Privacy Laws or Third Party Protection								
Japan	56	0.983	0.530	.0221	0.978	59.0	2	< 0.001
Korea	14	0.479	0.057	-0.036	0.831	1.9	2	0.167
Russian Federation	24	0.994	0.870	0.627	0.996	161.0	2	< 0.001
USA	1586	0.999	0.686	0.370	0.989	988.3	2	< 0.001

Table 3-6: Intraclass Correlation Between All Purpose Related P3P Tags (12 total tags)

				95% confidence interval		F-Test		
Country	Number Of websites	Cronbach's Alpha	Intraclass Correlation	lower	Upper	Value	df	*p*
Nations With General Privacy Laws and Third Party Protection								
Australia	25	0.911	0.223	0.110	0.472	11.3	11	< 0.001
Belgium	14	0.944	0.552	0.357	0.790	17.9	11	< 0.001
Canada	91	0.965	0.212	0.114	0.443	28.8	11	< 0.001
Denmark	22	0.895	0.209	0.099	0.455	9.5	11	< 0.001
France	40	0.956	0.331	0.189	0.596	22.7	11	< 0.001
Germany	80	0.977	0.305	0.175	0.563	42.8	11	< 0.001
Netherlands	63	0.967	0.231	0.125	0.472	30.7	11	< 0.001
New Zealand	9	0.951	0.661	0.462	0.857	20.5	11	< 0.001
Spain	15	0.934	0.450	0.266	0.715	15.1	11	< 0.001
Sweden	15	0.922	0.328	0.168	0.606	12.8	11	< 0.001
United Kingdom	176	0.988	0.275	0.158	0.524	83.8	11	< 0.001
Nations Without General Privacy Laws or Third Party Protection								
Japan	56	0.967	0.260	0.143	0.511	30.8	11	< 0.001
Korea	14	0.608	0.067	0.010	0.232	2.5	11	0.006
Russian Federation	24	0.995	0.885	0.788	0.957	184.8	11	< 0.001
USA	1586	0.996	0.257	0.147	0.500	238.6	11	< 0.001

Table 3-7: Intraclass Correlation Between All Category Related P3P Tags (17 total tags)

				95% confidence interval		F-Test		
Country	Number Of websites	Cronbach's Alpha	Intraclass Correlation	lower	Upper	Value	df	*p*
Nations With General Privacy Laws and Third Party Protection								
Australia	25	0.845	0.131	0.063	0.282	6.5	16	< 0.001
Belgium	14	0.853	0.273	0.146	0.493	6.8	16	< 0.001
Canada	91	0.960	0.149	0.085	0.294	24.8	16	< 0.001
Denmark	22	0.813	0.078	0.032	0.188	5.3	16	< 0.001
France	40	0.959	0.311	0.192	0.520	24.1	16	< 0.001
Germany	80	0.946	0.156	0.088	0.307	18.6	16	< 0.001
Netherlands	63	0.960	0.236	0.141	0.424	24.9	16	< 0.001
New Zealand	9	0.305	0.033	-0.019	0.158	1.4	16	0.133
Spain	15	0.820	0.223	0.111	0.432	5.6	16	< 0.001
Sweden	15	0.856	0.154	0.068	0.330	16.0	16	< 0.001
United Kingdom	176	0.982	0.192	0.115	0.358	55.6	16	< 0.001
Nations Without General Privacy Laws or Third Party Protection								
Japan	56	0.969	0.247	0.147	0.440	32.0	16	< 0.001
Korea	14	0.737	0.120	0.046	0.282	3.7	16	< 0.001
Russian Federation	24	0.996	0.888	0.810	0.949	225.6	16	< 0.001
USA	1586	0.993	0.176	0.105	0.332	148.1	16	< 0.001

Table 3-8: Intraclass Correlation Between All Retention Related P3P Tags (5 total tags)

				95% confidence interval		F-Test		
Country	Number Of websites	Cronbach's Alpha	Intraclass Correlation	Lower	Upper	Value	df	*p*
Nations With General Privacy Laws and Third Party Protection								
Australia	25	0.928	0.369	0.146	0.838	13.9	4	< 0.001
Belgium	14	0.938	0.551	0.261	0.916	16.1	4	< 0.001
Canada	91	0.951	0.202	0.075	0.686	20.4	4	< 0.001
Denmark	22	0.796	0.174	0.035	0.682	4.9	4	0.001
France	40	0.927	0.265	0.097	0.762	13.6	4	< 0.001
Germany	80	0.869	0.084	0.023	0.461	7.6	4	< 0.001
Netherlands	63	0.972	0.373	0.166	0.835	35.4	4	< 0.001
New Zealand	9	0.892	0.535	0.207	0.915	9.3	4	< 0.001
Spain	15	0.929	0.504	0.223	0.900	14.1	4	< 0.001
Sweden	15	0.785	0.223	0.041	0.749	4.6	4	< 0.003
United Kingdom	176	0.977	0.220	0.088	0.705	44.1	4	< 0.001
Nations Without General Privacy Laws or Third Party Protection								
Japan	56	0.939	0.253	0.095	0.748	16.4	4	< 0.001
Korea	14	0.421	0.054	-0.035	0.513	1.7	4	0.158
Russian Federation	24	0.987	0.789	0.556	0.969	76.2	4	< 0.001
USA	1586	0.994	0.270	0.116	0.755	165.8	4	< 0.001

Table 3-9: Intraclass Correlation Between All Recipient Related P3P Tags (6 total tags)

				95% confidence interval		F-Test		
Country	Number Of websites	Cronbach's Alpha	Intraclass Correlation	lower	upper	Value	df	*p*
Nations With General Privacy Laws and Third Party Protection								
Australia	25	0.971	0.555	0.309	0.886	34.0	5	< 0.001
Belgium	14	0.990	0.879	0.722	0.978	102.3	5	< 0.001
Canada	91	0.994	0.639	0.404	0.915	179.0	5	< 0.001
Denmark	22	0.973	0.653	0.401	0.921	37.3	5	< 0.001
France	40	0.994	0.807	0.612	0.962	169.5	5	< 0.001
Germany	80	0.996	0.744	0.527	0.946	243.4	5	< 0.001
Netherlands	58	0.995	0.758	0.544	0.950	208.2	5	< 0.001
New Zealand	9	0.987	0.896	0.747	0.982	78.4	5	< 0.001
Spain	15	0.995	0.936	0.842	0.989	220.0	5	< 0.001
Sweden	15	0.9910	0.885	0.737	0.979	116.8	5	< 0.001
United Kingdom	176	0.996	0.556	0.326	0.883	229.1	5	< 0.001
Nations Without General Privacy Laws or Third Party Protection								
Japan	56	0.987	0.457	0.238	0.837	74.15	5	< 0.001
Korea	14	0.858	0.266	0.086	0.716	7.0	5	< 0.001
Russian Federation	24	0.994	0.873	0.719	0.977	180.4	5	< 0.001
USA	1586	0.998	0.534	0.308	0.874	644.9	5	< 0.001

3.5.2.1 Evidence for hypotheses 6a and 6b

Through analysis of Table 3-3, no support for hypothesis 6a is found for any nation. This is particularly surprising since the European Union's Data Protection Directive [43] contains provisions concerning access privileges. For example, Article 41 requires that individuals be provided with access to information collected about themselves in order to ensure its accuracy [43]. This lack of consistent practice may be partly due to a lack of adherence. For instance, 20 websites from the Netherlands (32%), 26 from the United Kingdom (15%), 7 from Denmark (32%), 9 from France (23%), and 13 from Germany (16%), stated that they provided no access to selected information, which potentially violates Article 41. This potential lack of adherence to the legal framework is also observed in 15 Canadian websites (18%) who offered no access to selected information; an apparent contradiction of Principle 9 of the Personal Information Protection and Electronic Documents Act [44]. A similar potential lack of adherence has recently been identified by Bowie and Jamal [102] in UK websites. Their analysis found that many UK websites did not disclose the usage of cookies on a website; a potential violation of the European Union Data Directive. Our findings, if substantiated, provide further support for Bowie and Jamal's recommendation that formal government regulation should be resisted until it is understood why companies appear to not adhere to the currently enacted legislation in both the European Union and Canada.

While no support for Hypothesis 6a was found, Hypothesis 6b could only be rejected for the surveyed Russian websites. This apparent standardization of practice cannot be considered privacy protection; since 23 out of 24 websites provide individuals with no access to selected information. The OECD privacy principle of Individual Participation [103] requires that individuals have access to information about them, the right to challenge any information held, and to have that information updated if the challenge is successful. While the OECD privacy principles are not mandatory guidelines and Russia is not a signatory, they are a generally agreed upon framework for privacy protection supported by 30 nations.

3.5.2.2 Evidence for hypotheses 7a and 7b

While the results in Table 3-4 support hypothesis 7a for the United Kingdom, France, Australia, and Spain, these nations constitute but a minority of nations surveyed that have implemented general privacy legislation and provide third party enforcement. The surveyed websites from these nations all tend to remedy their disputes through customer service. It is unknown why websites from other nations have not followed this practice, since customer service would appear to be the preferable means of resolving disputes when compared to independent arbitration, court decisions, or referencing applicable laws [33]. Hypothesis 7b could only be rejected for Russian websites, which also prefer to utilize customer service for resolving disputes.

3.5.2.3 Evidence for hypotheses 8a and 8b

The results of Table 3-5 provide support for hypothesis 8a only for the United Kingdom, France, Australia, and Canada. Websites from all four of these nations predominately indicate that they offer to correct the problem. This would appear to be the preferable remedy from a business perspective when compared to the other choices of either monetary compensation, or following legally mandated remedies [33]. We are at a loss to explain this observed lack of uniformity in other surveyed nations. Hypothesis 8b was only rejected for Russian websites. It is surprising that more nations did not follow Russia's lead and prefer to fix the error rather than refer to laws or paying compensation.

3.5.2.4 Evidence for hypotheses 9a and 9b

The results of Table 3-6 indicate that there is no support for hypothesis 9a from any surveyed nation. This finding is especially surprising for France since they implement a strict licensing bureau [48]. The implications of these findings are significant for Internet users who do not read privacy policies [28]; no standard of practice whatsoever has evolved in the collection of non-transactional data. While legislative action may restrict unlawful actions of websites, lawful but undesirable actions may well be rampant. Clearly, consumer pressure for a standard of data-collection practice either does not exist, or is as yet ineffectual.

Hypothesis 9b can only be rejected for Russian websites. Further analysis of the agreement between Russian websites indicates that they use information for website administration, completion of requested activity, further development of the website, analysis of patrons in an anonymized fashion, and to make decisions about patrons in an anonymous fashion. While these results are not highly informative, it is unknown why Russian websites did not undertake more invasive analysis such as profiling of individuals, since few restraints appear to exist.

3.5.2.5 Evidence for hypothesis 10a

Hypothesis 10a can only be rejected for Russian websites given the results of Table 3-7. The observed Russian websites generally collected information relating to personal identification, characteristics of a patron's computer, site navigation and system state. The similarity in both types of information collected and purposes for its collection (outside of completing the current transaction) is unique in our survey. One possible explanation is that the surveyed Russian websites are related either through ownership or services offered. However, we have been unable to confirm or refute this possibility.

3.5.2.6 Evidence for hypothesis 10b

The results of Table 3-8 only allow for the rejection of hypothesis 10b for Russian websites. We summarize our data for Hypothesis 10b in Table 3-10, which outlines the retention periods used by surveyed websites. While general agreement was not expected, the number of websites who say they store information indefinitely is disturbing. In fact, the indefinite storage of information by 40 of 86 surveyed Canadian websites potentially

Table 3-10: Retention Periods Specified by Surveyed Websites In November 2005

	No retention	Stated purpose	Legal requirement	Business practices	Indefinitely
Australia	0 (0.0%)	2 (8%)	1 (4%)	6 (24%)	16 (64%)
Belgium	0 (0.0%)	0 (0.0%)	2 (14%)	4 (28%)	12 (85%)
Canada	7 (8%)	8 (9%)	10 (11%)	37 (41%)	45 (49%)
Denmark	5 (23%)	1 (5%)	0 (0.0%)	8 (36%)	10 (45%)
France	1 (3%)	9 (23%)	2 (5%)	18 (45%)	23 (58%)
Germany	12 (15.0%)	25 (31.3%)	6 (7.5%)	24 (30.0%)	33 (41.3%)
Netherlands	6 (10.3%)	10 (17.2%)	2 (3.4%)	22 (37.9%)	47 (81.0%)
New Zealand	0 (0.0%)	7 (78%)	0 (0.0%)	1 (11.1%)	1 (11.1%)
Spain	2 (13.3%)	0 (0.0%)	0 (0.0%)	4 (26.7%)	12 (80.0%)
Sweden	3 (20.0%)	1 (6.7%)	0 (0.0%)	7 (46.7%)	8 (53.3%)
U.K.	18 (10.2%)	29 (16.5%)	16 (9.1%)	74 (42.0%)	101 (57.4%)
Japan	5 (8.9%)	4 (7.1%)	1 (1.8%)	16 (28.6%)	30 (53.6%)
Korea	0 (0.0%)	0 (0.0%)	1 (7.1%)	3 (21.4%)	3 (21.4%)
Russia	0 (0.0%)	0 (0.0%)	0 (0.0%)	21 (87.5%)	2 (8.3%)
U.S.	125 (7.9%)	204 (12.7%)	85 (5.4%)	605 (38.1%)	880 (55.5%)

Note: Rows can sum to more than a nation's total number of websites since websites may utilize different storage lengths for different types of data violates the privacy principles outlined by the Model Code for the Protection of Personal Information [104] and appears inconsistent with Principle 5 governing retention periods in the Personal Information Protection and Electronic Documents Act [44]. Under certain circumstances, this form of indefinite retention may be appropriate as in government websites requiring payment information for old age security payments. However, under most circumstances, information is collected to satisfy short/medium term business practices that have a finite time frame. Under these circumstances, indefinite retention of information implies *permanent* vulnerability to security breaches within the data recipients, or privacy invasions by the data recipients.

3.5.2.7 Evidence for hypothesis 10c

From the results of Table 3-9, Hypothesis 10c was rejected for the populations of websites from Belgium, Spain, Sweden, France, New Zealand, and Russia. Websites from these nations (except Russia) all indicate that information will not be shared outside of the company. This agreement is unexpected since it would appear reasonable to hypothesize that many websites in the Alexa 100,000 list would be E-Commerce sites. Further it would also appear sensible that many of these E-Commerce companies would use third parties to ship their products. While unforeseen, this finding indicates that information retained by these websites may be relatively secure since information sharing, and by extension usage, would be restricted. In contrast, Russian websites generally share information with unrelated third parties who do not follow similar privacy practices.

3.6 Exploratory Analysis of Differences Between Cultures

In Section 3.5.1, empirical evidence has been identified supporting the theories of Shaw [11] and Hsu and Kuo [12]. Their theories however, also suggest that differences should exist in the content of P3P documents since P3P documents describe a multitude of privacy sensitive decisions made by organizations. The results of Table 3-11 indicate that these differences do indeed exist. Table 3-11 analyzes the differences between High and Low-Context cultures (the nations analyzed in Section 3.5.1). A weighted chi-square test was again employed. Hofstede's scores are not utilized in this analysis due to a lack of sufficient rationale for a dichotomization of the scores, which is required by the weighted chi-square test.

Table 3-11: Differences in Privacy Sensitive Actions Between Websites From High and Low Context Nations. (Part 1)

Weighted Chi-Square	*p*		Cohen's d
Type of Data Collected			
Information about an individuals computer (Operating system)	0.062	0.803	0.344
Content of communications (Text of Message)	1.225	0.268	0.095
Demographic information (Age, Sex)	0.230	0.632	0.305
Financial information (credit history)	2.616	0.106	0.157
Government issued identifiers (Social insurance/security number)	2.059	0.151	0.553
Health information (Personal health records)	0.015	0.904	0.778
Interactive information (Search queries, access logs)	0.449	0.502	0.091
Location information (GPS position)	0.774	0.389	0.405
Website navigation information (the pages viewed)	2.719	0.099	0.075
Online contact (email)	9.134	0.003	0.929
Information which does not fit into these categories	0.051	0.821	0.057
Physical contact information (telephone number)	5.159	0.023	0.722
Political information (religious, political, professional associations)	0.015	0.904	0.487
Personal preferences (Favorite color)	0.747	0.388	0.158
Purchase history (credit card information)	2.394	0.122	0.395
State management (cookie information)	0.539	0.463	0.201
Unique identifiers (login name)	13.194	< 0.001	0.250
Navigation	0.184	0.668	0.120

Table 3-11: Differences in Privacy Sensitive Actions Between Websites from High and Low Context Nations. (Part 2)

Weighted Chi-Square	*P*		Cohen's d
Purpose for Data Collection			
Administration of website	1.301	0.254	0.063
Contacting individuals for marketing of services or products	18.344	< 0.001	0.545
Completion of current activities	5.621	0.018	0.380
Research and development of services or products	4.155	0.042	0.161
Historical Preservation	3.001	0.083	0.244
Analysis of individuals	1.493	0.262	0.096
Making decisions based upon an analysis of an individual	4.148	0.042	0.110
A purpose other that these categories	0.957	0.328	0.155
Anonymous analysis of people	2.499	0.114	0.036
Making decisions based upon anonymous profiles	1.493	0.222	0.145
Tailoring the website to users preferences	15.416	< 0.001	0.244
Undertaking telemarketing	5.862	0.015	0.292
Who is allowed to access data			
Delivery services	5.294	0.021	0.151
People not fitting into these categories	1.089	0.297	0.455
The information is publicly available	0.001	0.972	0.472
Legal entities following our practices	0.877	0.349	0.089
Unrelated third parties	2.563	0.109	0.479
Ourselves and/or agents acting on our behalf	5.780	0.016	0.187
Are you allowed to access information about yourself			
All identifiable information	0.111	0.739	0.115
Identified contact information and other identified information	6.304	0.012	0.260
Identified contact information	0.283	0.595	0.197
No access to any identified information	0.000	0.993	0.115
Website does not collect information	2..377	0.123	0.219
Access is given to certain other identified information	7.321	0.007	0.778
What dispute resolution methods are applicable			
Court	1.884	0.170	0.981
Independent arbitration	2.939	0.086	0.153
Appropriate Laws	8.543	0.003	0.054
Customer service	1.567	0.211	0.626
Remedies that will be offered			
Correct the mistake	0.578	0.447	0.326
Provide cash settlement	1.108	0.292	0.484
What the law stipulates	3.308	0.069	0.140
How long will information be retained for			
As long as business practices warrant	13.374	< 0.000	0.451
Indefinitely	4.156	0.041	0.194
As long as legally required	7.235	0.007	0.666
Data not retained	2.380	0.123	0.042
Retained until stated purpose completed	0.565	0.452	0.021

The results of Table 3-11 indicate that a number of statistically significant ($p < 0.05$) differences, with medium ($d \geq 0.5$) and large ($d \geq 0.8$) effect sizes exist between websites from High-Context and Low-Context nations. This finding has significant implications for individuals who utilize websites spanning this cultural dimension; common standards of practice have not evolved, and so these users should not develop expectations about the protection of their privacy.

Through further ad-hoc analysis of Table 3-11, we determined that Low-Context nations were significantly more likely to collect online ($p < 0.003$, $d = 0.929$) and physical ($p < 0.023$, $d = 0.722$) contact information. Low-Context cultures were also more likely to use collected information for contacting individuals for marketing of services or products ($p < 0.001$, $d = 0.545$). This finding suggests that the increased collection of contact information may be due to the apparent increase in direct marketing of products and services in Low-Context cultures. Taylor *et al.* [105] has identified that the Japanese have a statistically significant more negative attitude towards direct marketers than Americans. Taylor *et al.* [105] hypothesizes that direct marketing methods force direct marketers into undertaking Low-Context communication which may offend Japanese sensibilities. If substantiated, these results suggest that interacting with websites from High-Context cultures may pose less of a threat for some forms of privacy invasion such as direct marketing. Websites from Low-Context cultures were also more likely to retain information for a legally mandated period ($p < 0.007$, $d = 0.666$). This is likely a consequence of the increased prevalence of privacy laws within these cultures.

The most general result of Table 3-11 is that websites from High-Context cultures are *never* more likely than websites from Low-Context cultures to state that they undertake any of the privacy-sensitive actions covered by P3P documents. All significant differences with moderate-to-large operational effects are related to websites from Low-Context cultures being more likely to declare that action. This finding may be a result of a predisposition to avoid disputes in High-Context cultures. For instance, Gudynkunst *et al.* [79] propose that all High-Context cultures are communalistic and all Low-Context cultures are individualistic. This implies that High-Context cultures would generally emphasize stability and harmony in business relationships [106], while Low-Context cultures would generally emphasize individual benefits. This could imply that, in order to preserve the stability and harmony of the business relationship, websites from High-Context cultures may simply choose to avoid privacy-sensitive actions as much as possible. However, further investigation is needed to explain this difference.

3.7 Exploratory Analysis of Differences Within Nations

While the analysis of Section 3.5.2 indicates that few standard practices exist between websites of a nation, it is unclear whether this is a result of a chaotic environment or one composed of contiguous clusters of websites following similar practices. If it is found that a series of relatively contiguous clusters do indeed exist, it may be possible for users to develop expectations regarding the actions of websites based upon the cluster they belong to. We propose to treat these clusters as latent categories, and to employ a variant of exploratory factor analysis (EFA) known as exploratory latent class analysis (ELCA) to identify these latent variables.

Factor analysis techniques are often employed when researchers whish to identify relationships between variables (latent factors). These procedures allow researchers to reduce the dimensionality of the system as well as develop insight into the dynamics of complex systems. When researchers have few (if any) expectations regarding the number or nature of the latent factors, EFA techniques are appropriate [107]. However, standard EFA techniques are not applicable to our data, as they require both observed and latent variables to be continuous in nature (not categorical). Tetrachoric correlation [108] and Item Response Theory [108] have been proposed as solutions to the dilemma of applying factor analysis techniques to data sets composed of observed categorical responses. By applying these methods, dichotomous responses can be mapped onto continuous latent variables; however, our interest is in discovering *classes* of websites with *homogenous* behavior; thus, we seek to map dichotomous responses onto a categorical latent variable which could allow for the possible development of a typology.

ELCA [70, 109, 110] is analogous to exploratory factor analysis when observed and latent variables are categorical. Through ELCA, observed response patterns, composed of nominal or ordinal measurements, are mapped onto a categorical latent variable whose levels represent the various latent classes. This mapping of observed responses onto a latent classification variable is analogous to exploratory cluster analysis and has been widely employed in the social sciences as a means of developing and testing typologies [110].

ELCA techniques utilize cross tabulation tables to represent the distribution of response patterns and rely upon latent class models to classify the response patterns into latent classes. The latent class models are composed of Latent Class and Conditional probabilities. Latent Class probabilities predict the probability that a response pattern belongs to a certain class given the value of one of the variables within the pattern. These probabilities allow researchers to determine the number and size of the resulting latent classes. The conditional probabilities represent the probability that an individual variable has a certain value, given that it belongs to a certain class. These probabilities can be used to determine the characteristics of individual latent classes. The model is able to predict the individual cell proportions for the cross tabulation table by summing the products of the relevant conditional probabilities and the latent class probabilities. This ability allows the latent class model to be tested against the observed data set to determine its fit. If statistically significant deviations exist between the predicted cross tabulation table and the observed table, the model must be rejected.

The generation of the conditional and latent class probabilities will use the expectation maximization (EM) procedure identified by Dempster *et al.* [111]. In their procedure, initial values for the conditional and latent class probabilities (which may be randomly selected) are continually improved through an iterative procedure whereby current estimations are compared against observed cell frequencies.

Due to the large number of observed dichotomous variables (53), the number of potential response patterns is very large (2^{53}). Since the ELCA model requires the usage of cross tabulation tables [110], these tables will naturally be sparse. However, the existence of sparse tables violate the assumption inherent with the most popular method of assessing model fit, the likelihood ratio Chi-Square goodness-of-fit test [110], since the test assumes that cell frequencies within the table all exceed a threshold (usually 5-10) [72]. Fitness testing for data such as ours is usually performed using resampling approaches [112, 113] or heuristics [110, 114].

Resampling methods are usually preferred over heuristics for determining model fit. Through these methods, pseudo-random populations are generated by recombining observations. Once generated, an estimator, such as the likelihood ratio Chi-Square goodness-of-fit test, is applied to the dataset. When this process is repeated a large number of times, a frequency distribution is generated for the chosen estimator thereby allowing researchers to determine whether the observed populations exhibit a statistically significant deviation from the null hypothesis. This approach avoids the problems associated with applying estimators on populations which violate the assumptions of models used to interpret the statistically significance indicated by the estimator.

In this analysis, we will employ the bootstrap resampling method since they are well suited for model testing [113]. In bootstrap resampling, pseudo-random populations are generated by sampling the observed dataset with replacement. The estimator employed will be the likelihood ratio Chi-Square goodness-of-fit test which is the standard method for assessing model fit. If the test statistic for the original dataset lies outside the 95% confidence interval of the distribution of statistic values for the pseudo-populations, we will reject the proposed model. Since exploratory latent class analysis using resampling techniques is dependent upon the sizes of the observed populations, its use on populations other than the United States (1,526 websites) is questionable. For this reason, resampling methods were only employed for this population.

Since ELCA does not require the researcher to hypothesize plausible latent structures, ELCA proceeds in a step wise manner whereby the population under analysis is first tested against a one latent class model. If a one latent class model cannot be rejected, then the only conclusion which can be reached is that the observed variables are not inter-related. If evidence exists allowing for the rejection of a one latent class model ($p < 0.05$) then a two latent class model is tested. This process continues until a model is found which cannot be rejected.

When we apply this approach to websites from the United States using the LCAP [115] ELCA tool, we found that a one latent class model must be rejected for the population of U.S. websites (p < 0.001) (Table 3-12). The results of Table 3-12 do suggest that a two latent class model provides sufficient fit to the observed dataset (p < 1.000). However, through an inspection of both the latent class and conditional probabilities, it was determined that the two latent class model classified websites based upon the number of P3P tags used in the document. Thus, websites which provided verbose P3P documents were categorized into one latent class while websites with relatively sparse P3P documents were categorized into the other latent class.

Since resampling methods yielded little insight which could be used for the development of non-trivial typologies, we now turn to heuristics to judge model fit. Caution is however urged in the interpretation of these results unless further evidence is found indicating these results to be statistically significant, or they are found to be significant through meta-analysis techniques in further studies [65]. Two popular heuristics utilized in ELCA are the Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) [114]. While some ELCA is completed through the use of only one heuristic, this practice is not recommended [114]. However, little guidance exists for how one should handle discrepancies between heuristics when two or more are utilized. Often the Scree test [116] is used as a mechanism of allowing researchers to visually determine where

Table 3-12: Results of Exploratory Latent Class Analysis

Number of Latent Classes	Total number of populations	Log likelihood Chi-Square	p	decision
1	1000	136938.516	< 0.001	reject
2	1000	44354.322	1.000	accept

significant changes have occurred when moving from one latent model to another. The Scree test relies on the identification of a "leveling off" point that is assumed to indicate the point where the addition of further factors (or classes) leads to minimal model improvement.

The results of an application of the Scree test to Figures 3-1 to 3-15 indicates that a two latent class model is appropriate for the US, UK, Sweden, Japan, and Canada. Due to the lack of an apparent leveling off point in the remaining populations, no conclusion may be made. The results of a visual inspection of the conditional probabilities in the aforementioned latent class models indicate that the clustering differentiates between verbose (many P3P tags) and sparse (few P3P tags) P3P documents. This dichotomization, which was previously identified for U.S. websites using resampling methods is of little practical value since it would appear natural to expect some P3P documents to use more P3P tags than others would. This inability to identify contiguous clusters of websites suggests that users cannot even create behavioral expectations for specific classes of websites within a single nation. This indicates that the Internet today appears to be a chaotic environment, with no standards of practice for privacy-sensitive actions. If supported, this would mean that users who hold behavioral expectations (outside of those gained from direct experience with a website) for privacy in the Internet environment are merely deluding themselves.

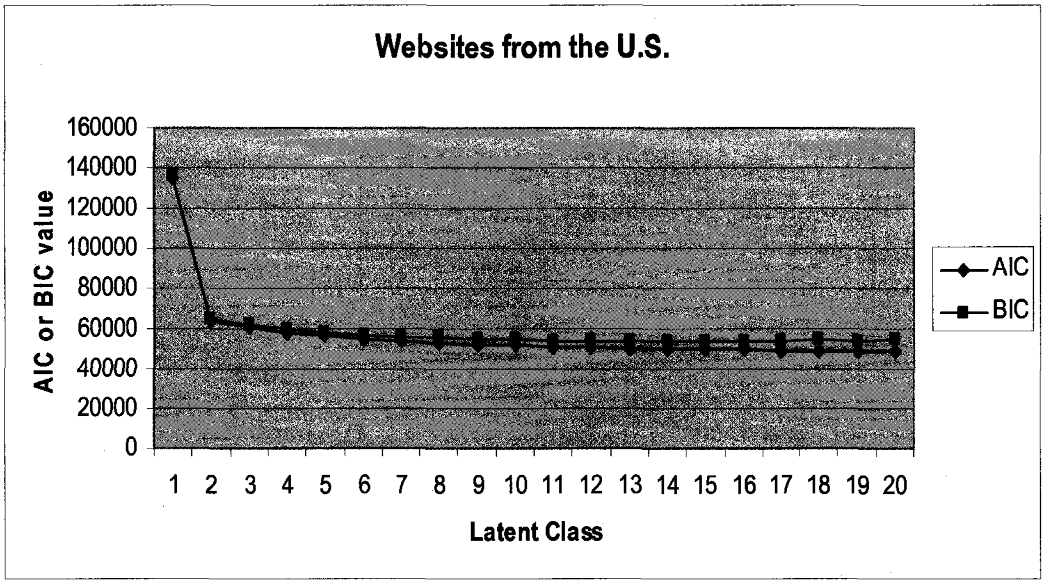

Figure 3-1: AIC/BIC Results for US Websites

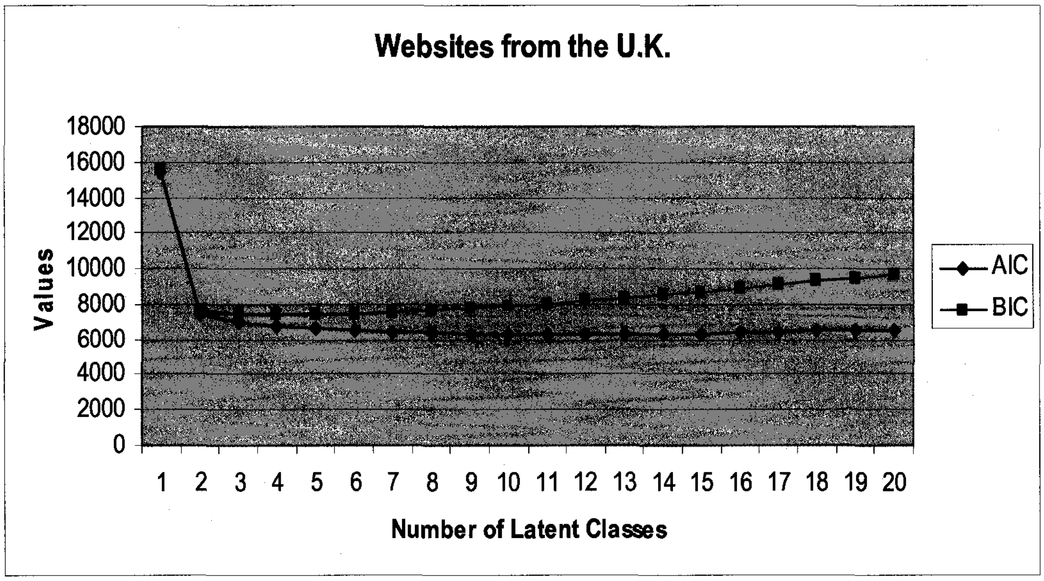

Figure 3-2: AIC/BIC Results for U.K. Websites

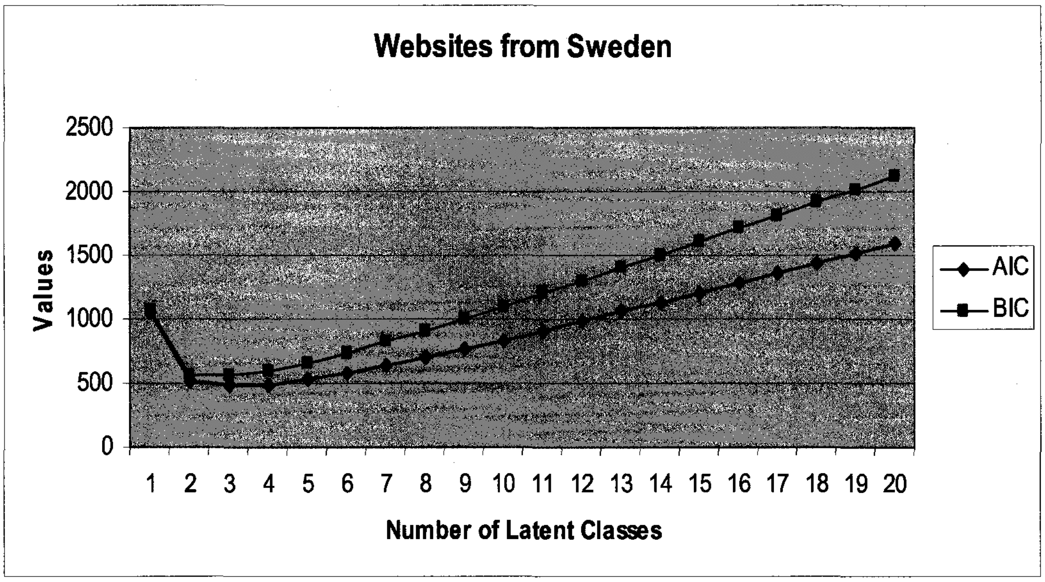

Figure 3-3: AIC/BIC Results for Swedish Websites

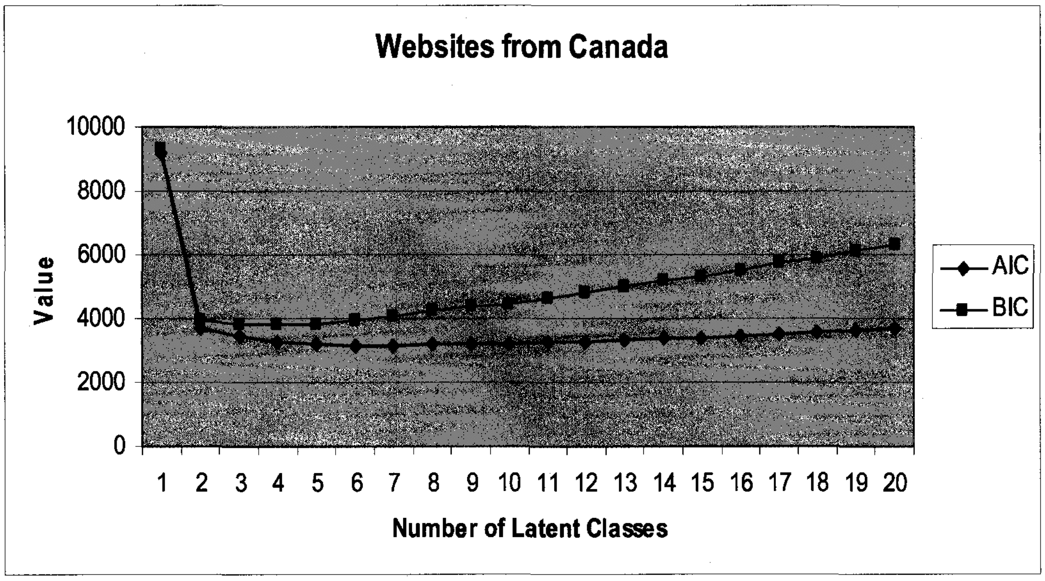

Figure 3-4: AIC/BIC Results for Canadian Websites

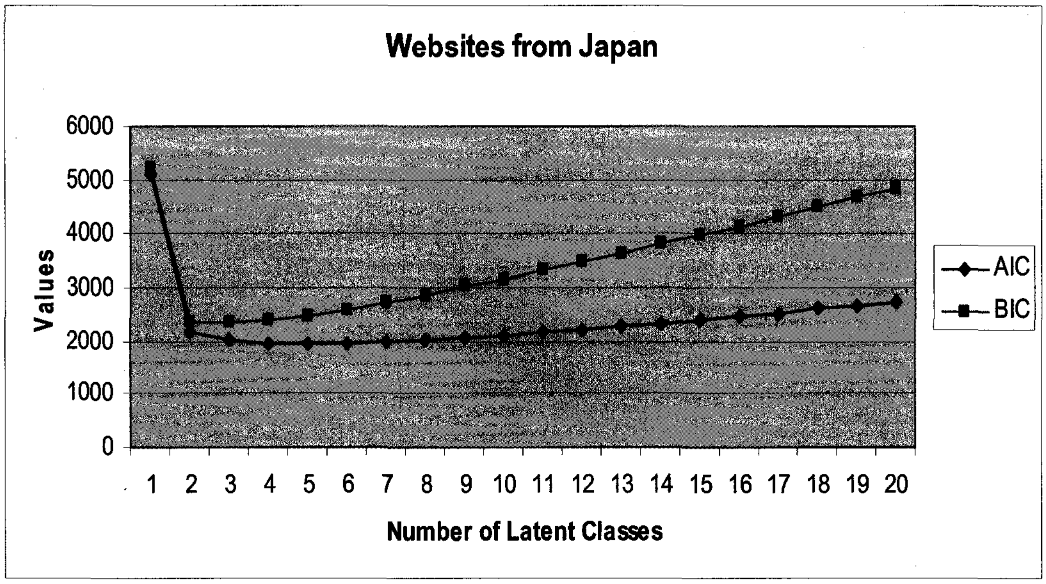

Figure 3-5: AIC/BIC Results for Japanese Websites

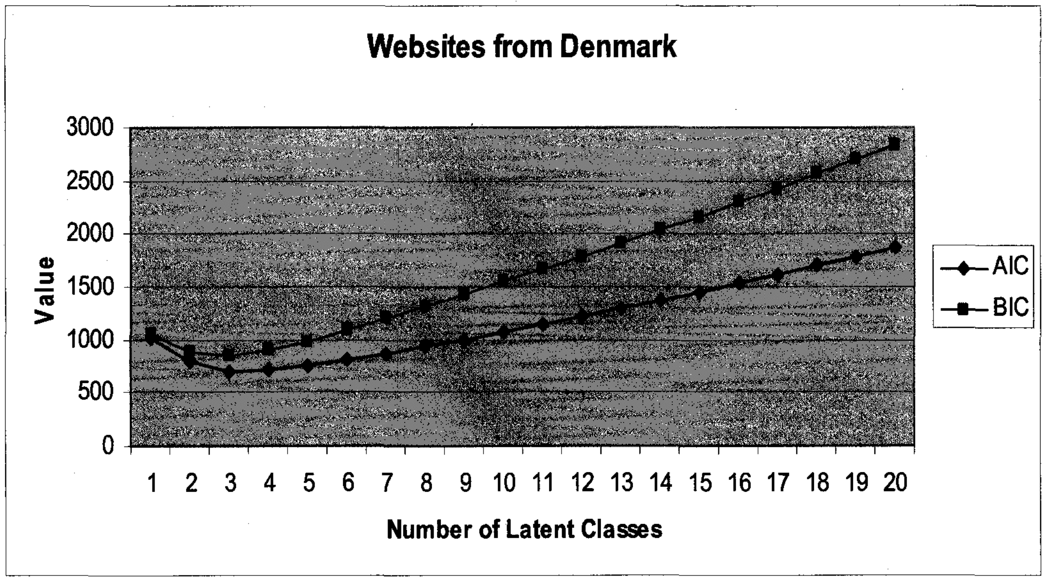

Figure 3-6: AIC/BIC Results for Danish Websites

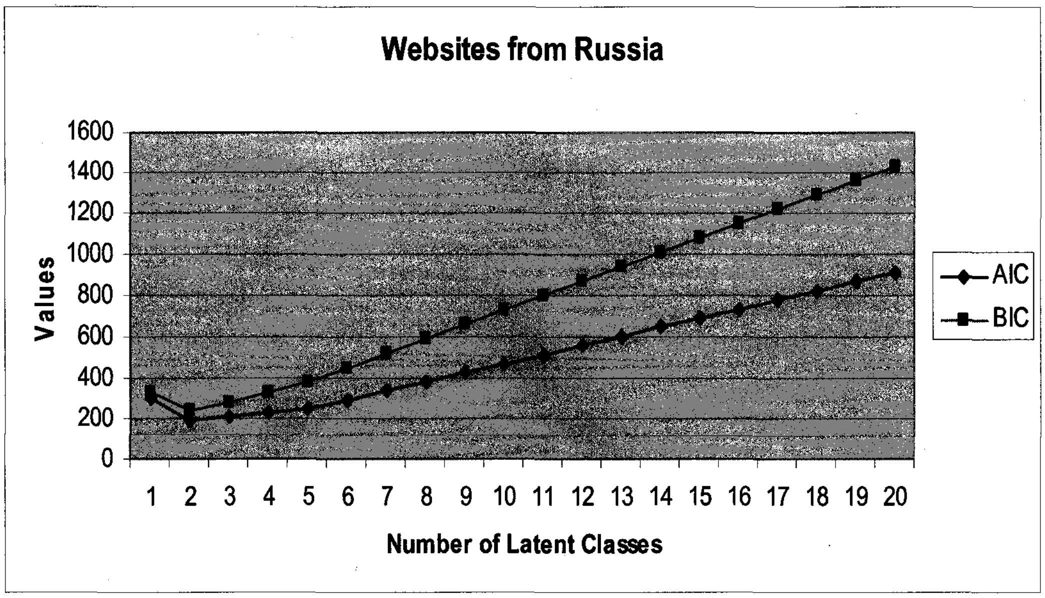

Figure 3-7: AIC/BIC Results for Russian Websites

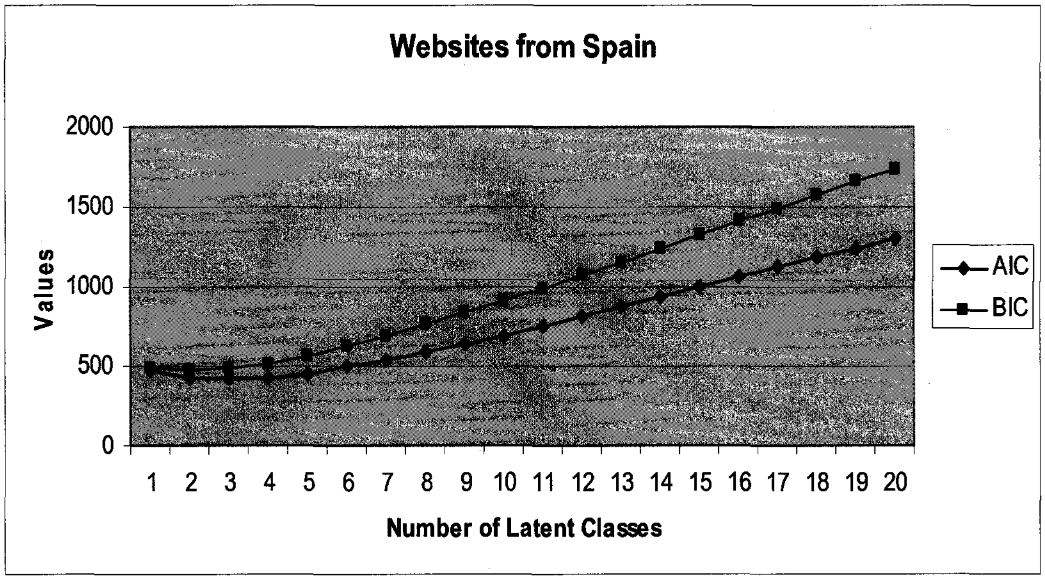

Figure 3-8: AIC/BIC Results for Spanish Websites

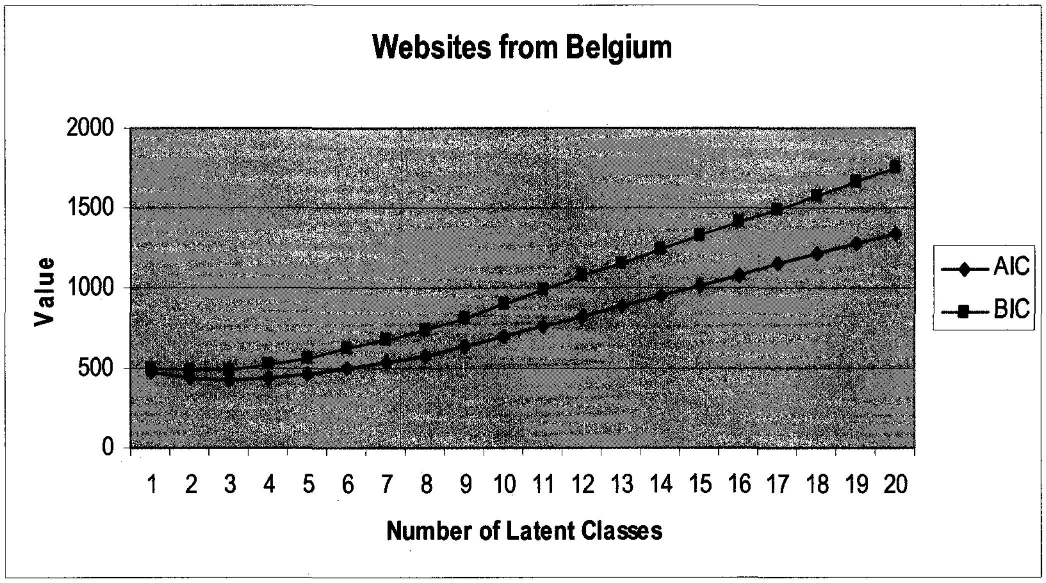

Figure 3-9: AIC/BIC Results for Belgian Websites

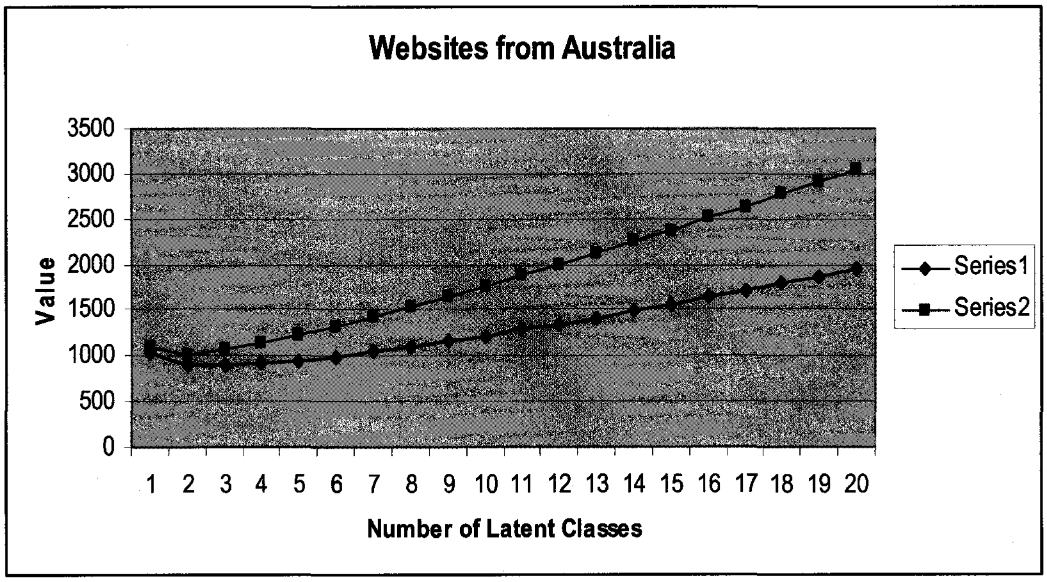

Figure 3-10: AIC/BIC Results for Australian Websites

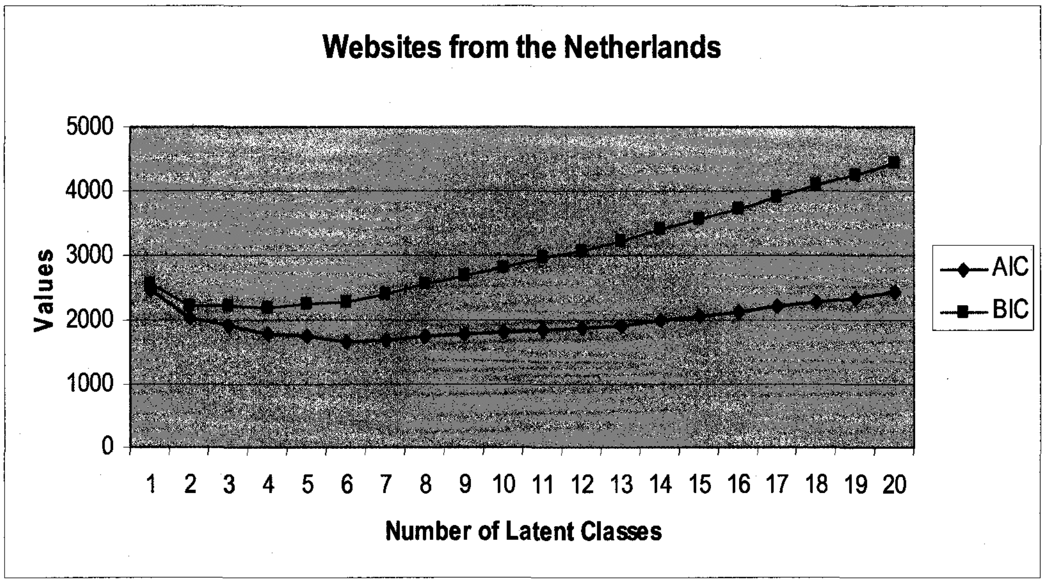

Figure 3-11: AIC/BIC Results for Dutch Websites

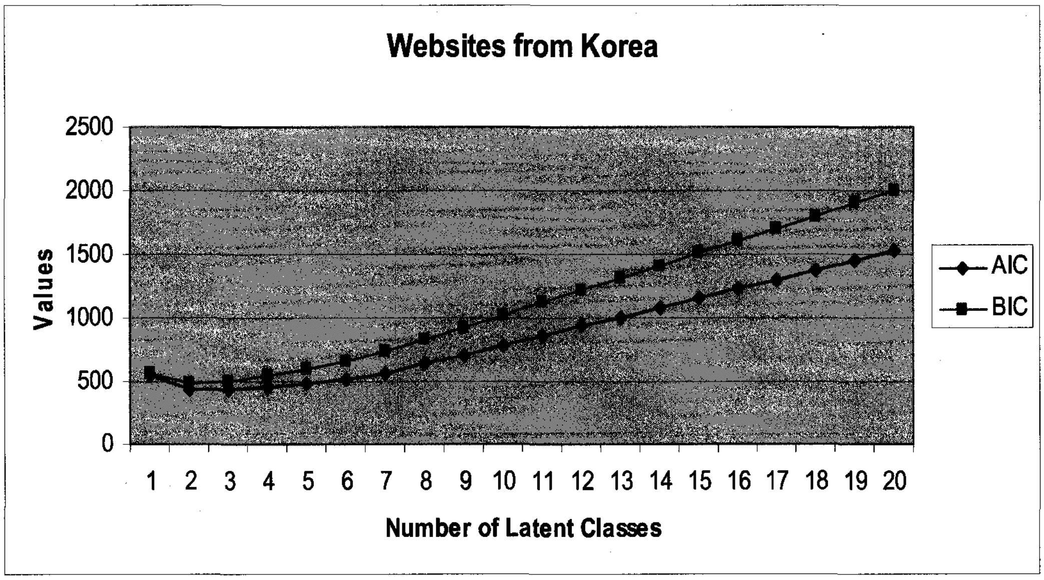

Figure 3-12: AIC/BIC Results for Korean Websites

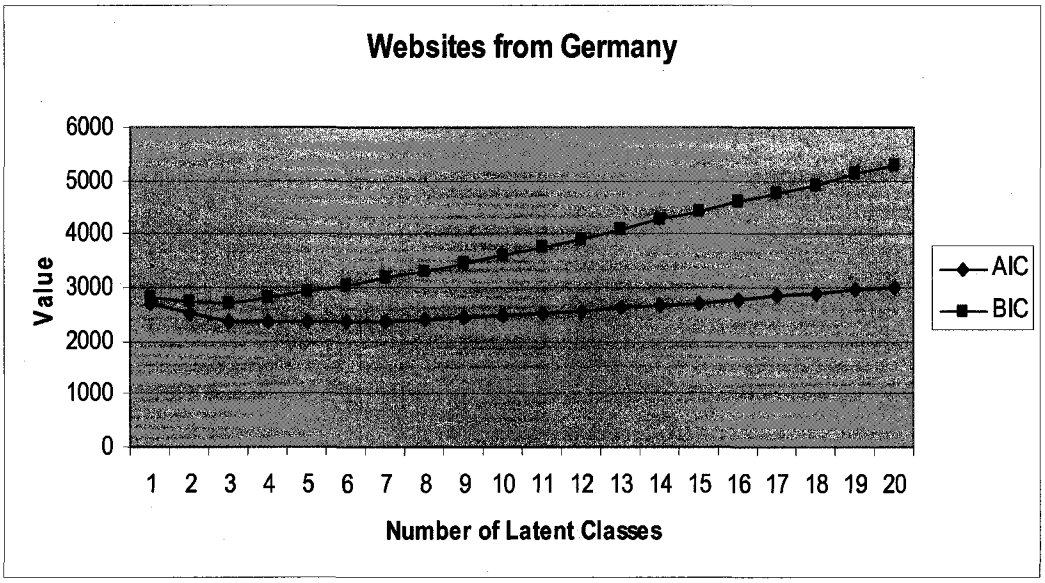

Figure 3-13: AIC/BIC Results for German Websites

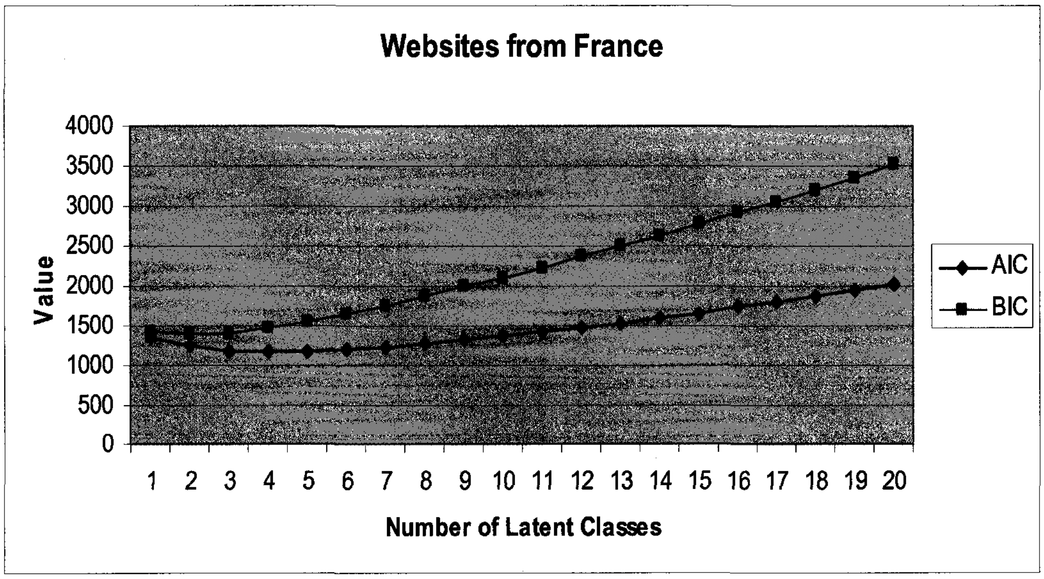

Figure 3-14: AIC/BIC Results for French Websites

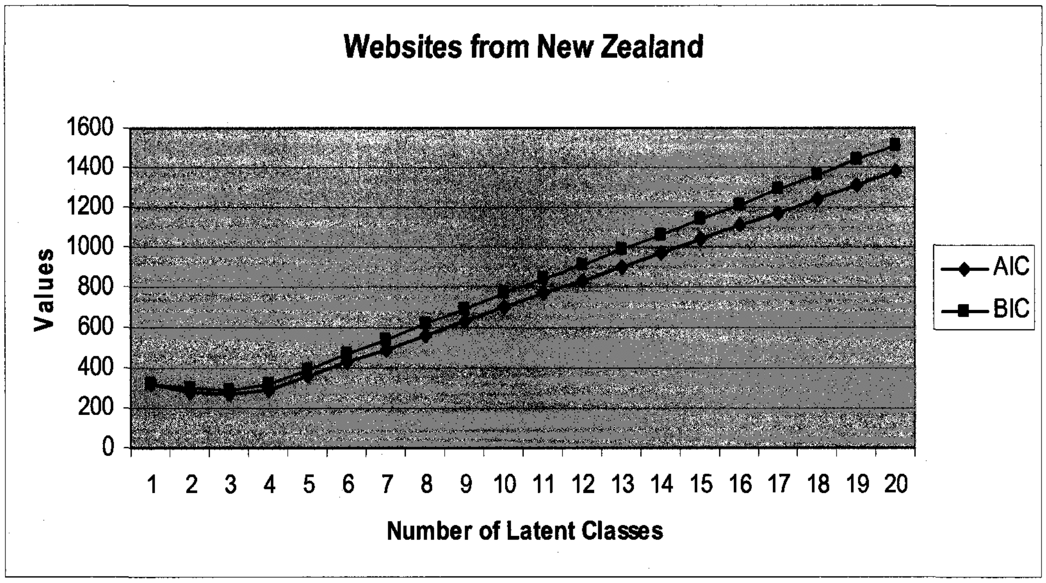

Figure 3-15: AIC/BIC Results for New Zealand Websites

3.8 Conclusion

In general, the findings of this survey suggest that Internet computing researchers need to begin to understand privacy not just from technological and legal perspectives, but also from a social perspective. The results of Section 3.5.1 indicate that the adoption of P3P varies across cultures, and empirically supports the findings of Shaw [11] and Hsu and Kuo [12]. The results from Section 3.6 provide further evidence by showing that websites from Low-Context cultures generally collect and utilize more information than their High-Context counterparts. The results of Section 3.5.1 also suggest that assumptions that individuals will adopt a particular PET simply because they are concerned about their privacy are ill founded. These results have consequences for all future development of PETs and suggest that any technology that is intended to be an Internet wide solution must be designed in a culturally sensitive manner.

The findings of Section 3.5.2 illustrate that many of the assumptions held by individuals are likely to be incorrect. Our results suggest that in many jurisdictions, assumptions regarding legal protections and actions of organizations analogous to those outlined by Turow *et al.* [1] are baseless due to the range of actions undertaken by websites. While this fact has often been taken for granted, it has never been empirically studied on a large scale. The actions of websites vary even where strict licensing bureaus are established such as in France [48]. The results of Section 3.7 furthers this argument by suggesting that groups of websites who follow similar practices are not easily identifiable. These results would appear to imply that without a novel approach to influencing websites internationally, calls for standardization [117] and alignment [2] are unlikely to be realized.

A limiting factor of our analysis, by necessity, is the usage of P3P documents as the primary information source. The limited adoption rate of these policies restricts the generalizability of the results and should be taken into account in the interpretation of the above results. The implications of this research illustrate several potential avenues for future work. Evidence on how social factors have hindered or advanced the adoption and utilization of other privacy enhancing technologies is plainly needed. This information would be instrumental in developing usable and realistic privacy solutions. Further evidence on the effect legislation, and enforcement, mechanisms have upon the actual actions of Internet organizations is also required. For instance, Section 3.5.2 indicates several instances where it appears that websites are possibly not adhering to legislation. Finally, information is required which further describes the actual actions undertaken by websites, beyond just their self-reports. Without this knowledge, attempts to educate the general public would appear baseless.

3.9 Bibliography

[1] J. Turow, L. Feldman, and K. Meltzer, "Open To Exploitation: American Shoppers Online and Offline," University of Pennsylvania's Annenberg School for Communication 2005.

[2] J. B. Earp, A. I. Antón, L. Aiman-Smith, and W. H. Stufflebeam, "Examining Internet Privacy Policies Within the Context of User Privacy Values," *IEEE Transactions on Engineering Management,* vol. 52, pp. 227-237, May 2005.

[3] "A Matter of Trust: What Users Want From Web Sites," in *Consumer WebWatch*: Consumer WebWatch, 2002.

[4] H. H. Clark, *Using Language*. New York: Cambridge University Press, 1996.

[5] A. I. Antón, J. B. Earp, Q. He, W. Stufflebeam, D. Bolchini, and C. Jensen, "Financial Privacy Policies and the Need for Standardization," *IEEE security and privacy,* vol. 2, pp. 36-45, March-April 2004.

[6] D. H. Bracey, *Exploring Law and Culture*. Long Grove, Illinois: Waveland Press, Inc., 2006.

[7] R. C. Ellickson, "The Evolution of Social Norms: A Perspective from the Legal Academy," in *Social Norms*, M. Hechter and K.-D. Opp, Eds. New York: Russell Sage Foundation, 2001, pp. 35-75.

[8] C. Horne, "Sociological Perspectives on the Emergence of Norms," in *Social Norms*, M. Hechter and K.-D. Opp, Eds. New York: Russell Sage Foundation, 2001.

[9] S. Greenspan, D. Goldberg, D. Weimer, and A. Basso, "Interpersonal Trust and Common Ground in Electronically Mediated Communication," in *2000 ACM Conference on Computer Supported Cooperative Work*, Philadelphia, Pennsylvania, United States, 2000, pp. 251-260.

[10] S. E. Kaplan and R. J. Nieschwietz, "A Web Assurance Model of Trust for B2C E-Commerce," *International Journal of Accounting Information Systems,* vol. 4, pp. 95-114, June 2003 2003.

[11] T. R. Shaw, "The Moral Intensity of Privacy: An Empirical Study of Webmasters' Attitudes," *Journal of Business Ethics,* vol. 46, pp. 301-318, Sep 2003.

[12] M.-H. Hsu and F.-Y. Kuo, "The Effect of Organization-Based Self-Esteem and Deindividualism in Protecting Personal Information Privacy," *Journal of Business Ethics,* vol. 42, pp. 305-320, February 2003.

[13] R. Capurro, "Privacy. An Intercultural Perspective," *Ethics and Information Technology,* vol. 7, pp. 37-47, 2005.

[14] P. Kumaraguru and L. Cranor, "Privacy in India: Attitudes and Awareness," in *2005 Workshop on Privacy Enhancing Technologies*, Dubrovnik, Croatia, 2005.

[15] "Alexa Traffic Rankings." vol. 2006, 2006.

[16] "The Platform for Privacy Preferences (P3P) Project." vol. 2006, 2002.

[17] C. Jensen and C. Potts, "Privacy Policies as Decision-Making Tools: An Evaluation of Online Privacy Notices," in *CHI 2004*, Vienna, Austria, 2004.

[18] S. Lichtenstein, P. M. C. Swatman, and K. Babu, "Adding Value to Online Privacy for Consumers: Remedying Deficiencies in Online Privacy Policies with an Holistic Approach," in *36th Hawaii International Conference on System Sciences,*

Hawaii, 2003.

[19] "BBBonline." vol. 2006: Better Business Bureau, 2006.

[20] "Truste." vol. 2006: Truste.org, 2006.

[21] ScanAlert, "HackerSafe." vol. 2006: ScanAlert.

[22] "Web Trust." vol. 2006, 2006.

[23] L. Bruckner and M. Voss, "MozPETs - A Privacy Enhanced Web Browser," in *Third Annual Conference on Privacy, Security and Trust*, St. Andrews, New Brunswick, Canada, 2005.

[24] S. Gritzalis, "Enhancing Web Privacy and Anonymity In the Digital Era," *Information Management and Computer Security,* vol. 12, pp. 255-288, 2004.

[25] J. Goecks and E. D. Mynatt, "Social Approaches to End-User Privacy Management," in *Security and Usability Designing Secure Systems That People Can Use*, L. F. Cranor and S. Garfinkel, Eds. Beijing: O'Reilly, 2005.

[26] S. Garfinkel, *PGP: Pretty Good Privacy*. Sebastopol, California: O'Reilly & Associates, 1995.

[27] A. AcQuisti and J. Grossklags, "Privacy and Rationality in Individual Decision Making," *IEEE security and privacy,* vol. 3, pp. 26-33, Jan-Feb 2005.

[28] C. Jensen, C. Potts, and C. Jensen, "Privacy Practices of Internet Users: Self-Reports Versus Observed Behavior," *International Journal Human Computer Studies,* vol. 63, pp. 203-227, 2005.

[29] "WebTrust." vol. 2006: American Institute of Certified Public Accountants.

[30] "VeriSign." vol. 2006, 2006.

[31] T. Moores, "Do Consumers Understand the Role of Privacy Seals in E-

Commerce?," *Communications of the ACM,* vol. 48, pp. 86-91, March 2005 2005.

[32] A. Watt, *Beginning Regular Expressions*. Indianapolis Indiana: Wiley Publishing, Inc., 2005.

[33] "The Platform for Privacy Preferences 1.0 Specification." vol. 2006: World Wide Web Consortium (W3C), 2002.

[34] S. Byers, L. F. Cranor, and D. Kormann, "Automated Analysis of P3P-Enabled Web Sites," in *Fifth International Conference on Electronic Commerce*, Pittsburgh, 2003.

[35] L. F. Cranor, S. Byers, and D. Kormann, "An Analysis of P3P Deployment on Commercial, Government and Children's Web Sites as of May 2003," Federal Trade Commission, Trade Report May 2003 2003.

[36] I. Reay, P. Beatty, S. Dick, and J. Miller, "A Survey and Analysis of the P3P Protocol's Agents, Adoption, Maintenance, and Future," *Submitted for Review to IEEE Transactions on Dependable and Secure Computing,* 2006.

[37] S. Egelman and L. F. Cranor, "An Analysis of P3P-Enabled Web Sites among Top-20 Search Results," in *Eighth International Conference on Electronic Commerce*, Fredericton, New Brunswick, Canada, 2006.

[38] "US E-Government Act of 2002, Public Law 107-347-DEC. 17 2002," 2002.

[39] M. Henry, *International Privacy, Publicity & Personality Laws*. Markham, Ontario: Butterworths, 2001.

[40] E. Perkins and M. Markel, "Multinational Data-Privacy Laws: An Introduction for IT Managers," *IEEE Transactions on Professional Communication,* vol. 47, pp. 85-94, 2004.

[41] "First Report on the Implementation of the Data Protection Directive (95/24/EC)," Commission of the European Communities, 2003.

[42] "Privacy Initiatives." vol. 2006: Federal Trade Commission, 2006.

[43] "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data," in *23.11.1995*, 1995.

[44] "Personal Information Protection and Electronic Documents Act," Second ed, 2000.

[45] "Act on the Protection of Personal Information," 2003.

[46] "Canadian Privacy Commissioner." vol. 2006, 2006.

[47] "Information Commissioner's Office." vol. 2006: Information Commissioner, 2006.

[48] "Comission Nationale De L'Informatique Et Des Libertes." vol. 2006: Comission Nationale De L'Informatique Et Des Libertes, 2006.

[49] W. B. Gudynkunst, *Bridging Differences Effective Intergroup Communication*, 3rd ed. Thousand Oaks: Sage Publications, 1998.

[50] S. J. Milberg, H. J. Smith, and S. J. Burke, "Information Privacy: Corporate Management and National Regulation," *Organization Science,* vol. 11, pp. 35-57, January-February 2000.

[51] A. F. Westin, *Privacy and Freedom*. New York: Atheneum Publishers, 1967.

[52] S. Dahl, "Intercultural Research: The Current State of Knowledge," in *Middlesex University Discussion Paper*: Middlesex University, 2004.

[53] N. Singh, H. Zhao, and X. Hu, "Analyzing the Cultural Content of Web Sites, A

Cross-National Comparison of China, India, Japan, and US," *International Marketing Review,* vol. 22, pp. 129-146, 2003.

[54] G. Hofstede, "Cultural Dimensions." vol. 2006, 2006.

[55] E. T. Hall, *Beyond Culture.* New York: Doubleday, 1989.

[56] B. McSweeney, "Hofstede's Model of National Cultural Differences and their Consequences: A Triumph of Faith - A Failure of Analysis," *Human Relations,* vol. 55, pp. 89-118, January 2002.

[57] C. Ess and F. Sudweeks, "Culture and Computer-Mediated Communication: Toward New Understandings," *Journal of Computer-Mediated Communication,* vol. 11, 2005.

[58] "State Sues Major "Spyware" Distributor." vol. 2006: Office of New York State Attorney General Eliot Spitzer, 2006.

[59] "P3P Dashboard Report," Ernst and Young January 2003 2003.

[60] "P3P Validator." vol. 2006, 2002.

[61] "MySQL database." vol. 2006, 2005.

[62] "IP2Location." vol. 2006, 2006.

[63] C. Ahn, S.-H. Jung, and S.-H. Kang, "An Evaluation of Weighted Chi-Square Statistics for Clustered Binary Data," *Drug Information Journal,* vol. 37, pp. 91-99, 2003.

[64] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences (2nd Edition).* Hillsdale, NJ: Lawrence Earlbaum Associates, 1988.

[65] M. W. Lipsey and D. B. Wilson, *Practical Meta-Analysis* vol. 49. Thousand Oaks, California: Sage Publications, 2001.

[66] V. Hasselblad and L. V. Hedges, "Meta-Analysis of Screening and Diagnostic Tests," *Psychological Bulletin,* vol. 117, pp. 167-178, January 1995.

[67] D. C. Howell, *Statistical Methods for Psychology*, 5th ed. Pacific Grove, CA: Duxbury/Thomson Learning, 2002.

[68] L. J. Cronbach, "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika,* vol. 16, pp. 297 - 334, September 1951.

[69] J. C. Nunnally, *Psychometric Theory*, 2nd ed. New York: McGraw-Hill Book Company, 1978.

[70] E. B. Andersen, *The Statistical Analysis of Categorical Data.* New York: Springer-Verlag, 1990.

[71] L. G. Portney and M. P. Watkins, *Foundations of Clinical Research, Applications to Practice.* Norwalk, Connecticut: Appletop & Lange, 1993.

[72] D. J. Sheskin, *Handbook of Parametric and NonParametric Statistical Procedures*, 3rd ed. Boca Raton: Chapman and Hall/CRC, 2004.

[73] P. E. Shrout and J. L. Fleiss, "Intraclass Correlations: Uses in Assessing Rater Reliability " *Psychological Bulletin,* vol. 86, pp. 420-428, 1979.

[74] D. Garson, "Scales and Standard Measures." vol. 2006, 2006.

[75] J. C. Nunnally and I. H. Bernstien, *Psychometric Theory,* 3rd ed. New York: McGraw-Hill, Inc., 1994.

[76] S. D. Walter, M. Eliasziw, and A. Donner, "Sample size and optimal designs for reliability studies," *Statistics in Medicine,* vol. 17, pp. 101-110, January 1998.

[77] H. S. Bloom, "Minimum Detectable Effects," *Evaluation Review,* vol. 19, pp. 547-556, 1995.

[78] D. G. Bonett, "Sample Size Requirements for Estimating Intraclass Correlations with Desired Precision," *Statistics in Medicine,* vol. 21, pp. 1331-1335, May 2002.

[79] W. B. Gudynkunst, S. Ting-Toomey, and E. Chua, *Culture and Interpersonal Communication*. Newbury Park, California: Sage, 1988.

[80] S. Spiekermann, J. Grossklags, and B. Berendt, "E-Privacy in 2nd Generation E-Commerce: Privacy Preferences Versus Actual Behavior," in *3rd ACM conference on Electronic Commerce*, 2001.

[81] "Wet Bescherming Persoonsgegevens (Personal Data Protection Act)," 1999.

[82] "The Dutch Data Protection Authority." vol. 2006, 2006.

[83] "Law of December 8, 1992 on Privacy Protection in Relation to the Processing of Personal Data as Modified by the Law of December 11, 1998 Implementing Directive 95/46/EC," 1998.

[84] "Commission for the Protection of Privacy." vol. 2006, 2006.

[85] "Organic Law 15/1999 of 13 December on the Protection of Personal Data," in *23750*, 1999.

[86] "Agencia Espanola De Proteccion De Datos." vol. 2006, 2006.

[87] "Personal Data Act," in *204*, 1998.

[88] "Swedish Data Inspection Board." vol. 2006, 2006.

[89] "Data Protection Act 1998," in *29*, 1998.

[90] "Act of Processing Of Personal Data," 2000.

[91] "The Danish Data Protection Agency." vol. 2006, 2006.

[92] "Act n78-17 of 6 January 1978 on Data Processing, Data Files, and Individual

Liberties. Amended by the Act of 6 August 2004 relating to the protection of individuals with regard to the processing of personal data," in *n78-17*, 1978.

[93] "Federal Commissioner for Data Protection and Freedom of Information." vol. 2006, 2006.

[94] "Federal Data Protection Act," in *OJ EC no. L 281, p. 31 ff.*, 2003.

[95] "Privacy Act of 1988," in *119*, 1988.

[96] "Australian Government Office of the Privacy Commissioner." vol. 2006, 2006.

[97] "Privacy Act 1993," 1993.

[98] "Privacy Commissioner Te Matapono Matatapu." vol. 2006, 2006.

[99] "The Republic of Korea." vol. 2006: Privacy International.org, 2004.

[100] "PHR 2004 - The Russian Federation." vol. 2006, PrivacyInternational.org, Ed.: PrivacyInternational.org, 2004.

[101] M. T. S. Rajan, "The Past and Future of Privacy in Russia," *27 Review of Central and Easy European Law,* vol. 4, pp. 625-638, 2002.

[102] N. E. Bowie and K. Jamal, "Privacy Rights on the Internet: Self-Regulation or Government Regulation," *Business Ethics Quarterly,* vol. 16, July 2006.

[103] "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data." vol. 2006, 1980.

[104] "Model Code for the Protection of Personal Information." vol. 2006: Canadian Standards Association, 1996.

[105] C. R. Taylor, G. R. Franke, and M. L. Maynard, "Attitudes Toward Direct Marketing and Its Regulation: A Comparison of the United States and Japan," *Journal of Public Policy and Marketing,* vol. 19, pp. 228-237, 2000.

[106] D. A. Pitta, H.-G. Fung, and S. Isberg, "Ethical Issues Across Cultures: Managing the Differing Perspectives of China and the USA," *Journal of Consumer Marketing,* vol. 16, pp. 240-256, 1999.

[107] B. Thompson, *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*, 1st ed. Washington DC: American Psychological Association, 2004.

[108] R. J. Mislevy, "Recent Developments in the Factor Analysis of Categorical Variables," *Journal of Educational Statistics,* vol. 11, pp. 3-31, Spring 1986.

[109] A. Agresti, *Categorical Data Analysis*, 2nd ed. New York: Wiley-Interscience, 2002.

[110] A. L. McCutcheon, *Latent Class Analysis*. Newbury Park: Sage Publications, Inc, 1987.

[111] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological),* vol. 39, pp. 1-38, 1977.

[112] J. G. Dias and J. K. Vermunt, "Bootstrap Methods for Measuring Classification Uncertainty in Latent Class Analysis," in *Proceedings in Computational Statistics*, 2006.

[113] R. Langeheine, J. Pannekoek, and F. V. D. Pol, "Bootstrapping Goodness-of-Fit Measures in Categorical Data Analysis," *Sociological Methods and Research,* vol. 24, pp. 492-516, May 1996.

[114] J. Kuha, "AIC and BIC Comparisons of Assumptions and Performance," *Sociological Methods and Research,* vol. 33, pp. 188-229, November 2004.

[115] "LCAP." vol. 2006: Statistical Genetics Group, 2003.

[116] R. B. Cattell, "The Scree Test for the Number of Factors," *Multivariate Behavioral Research,* vol. 1, 245-276 1966.

[117] A. I. Antón, Q. He, and D. L. Baumer, "Inside JetBlue's Privacy Policy Violations," *IEEE security and privacy,* vol. 2, 2004.

A LARGE-SCALE EMPIRICAL STUDY OF ONLINE PRIVACY POLICIES: STATED ACTIONS VS. LEGAL OBLIGATIONS

4 Chapter 4 Introduction

Before interconnected technology, people were protected from many forms of privacy invasion since information collection, analysis, and usage were not cost effective. However, with the growth of Internet and Web computing technologies, cheap data storage, and effective data mining techniques, the information supplied by Web users can now be used for a range of novel actions previously unavailable to organizations. For instance, a recent report by the Canadian Internet Policy and Public Interest Clinic (CIPPIC) describes [1] how an entire industry has developed to supply organizations with datasets comprised of transaction histories and personal information aggregated across thousands of communities. The primary beneficiary of these aggregated data sets is the direct marketing industry, which uses this data for customer profiling. Profiling can benefit both parties in a transaction; services can be tailored to customers' needs and the total volume of advertising can be reduced and more effectively targeted. However, these datasets can also contain highly sensitive payment, health, and financial information. When used inappropriately or stolen, this information can be used for criminal purposes such as ID theft and credit card fraud. There is also a range of perfectly legal, but often unpalatable, uses for profiling. For example, unsolicited advertising, telemarketing and differential pricing are all practices that many consumers find annoying or even outrageous. For many people, such actions are a breach of their personal privacy, and therefore highly objectionable.

In an effort to alleviate the risks associated with information collection, aggregation and usage, a number of nations have proposed privacy protection laws that establish minimum standards for data collectors on how to responsibly use information. These principles often define the rights of individuals, the responsibilities of data collectors, and the methods for dispute resolution. Generally, these laws are enforced through ombudsmen (e.g. the Privacy Commissioner of Canada), or licensing bureaus (e.g. the CNIL in France). Recent reports, however, suggest that adherence to these programs may be questionable at best [2-4]. The current paper provides new evidence on website adherence to legal mandates by the Internet's most popular websites using Platform for Privacy Preferences (P3P) [5] documents posted on those sites. P3P documents are a machine readable version of a website's privacy policy, and thus provide the opportunity to perform an automated analysis of legal adherence that would otherwise be impractical to accomplish. The results of our survey of the Internet's most popular 100,000 websites suggest that adherence to legislation or programs across all domains is questionable at best and the effectiveness of legal mandates appear debatable.

The remainder of this paper is organized as follows: Section 4.1 overviews the current state of legislation, regulation, and enforcement mechanisms. Section 4.2 introduces the P3P protocol. Section 4.3 describes our survey methodology and data collection. Section 4.4 presents preliminary results and develops rule sets that test adherence to a jurisdiction's privacy principles. Section 4.5 presents and analyzes our results, and Section 4.6 provides our conclusions and a discussion of future work.

4.1 Legislation

Academia, business, and governments have proposed a range of solutions for solving the privacy and information security problems which currently plague the Internet. These solutions include the adoption of privacy enhancing technologies by Internet users, auditing and privacy seal programs for organizations, and the enactment of privacy legislation by national or state governments. While all three of these approaches have had limited success in solving the current online privacy problems, privacy legislation continues to be advocated for and actively considered by both national governments and international bodies such as APEC [6]. A unifying element of existing and proposed legislation is an attempt to protect several broadly agreed-upon privacy principles. These principles are often founded at least partially upon the OECD's privacy principles [7], which provide recommendations for reasonable information usage. Two major models of legislation have emerged: sectoral and comprehensive privacy legislation.

Sectoral laws have been implemented by a number of nations including the United States and South Korea. These laws govern the usage of sensitive information within a given industrial sector. For example, legislation in the USA governs websites offering financial services [8, 9] or services for children [10]. In contrast, comprehensive laws apply to all organizations within a legal jurisdiction. Examples of such laws are national implementations of the European Union's Data Protection Directive [11] and Canada's PIPEDA [12]. While these laws can create a uniform regulatory environment within a legal jurisdiction, conflict may occur when organizations need to transfer information between jurisdictions. For instance, Article 25 of the Data Protection Directive states that information can only be transferred to nations whose privacy laws are deemed to provide an 'adequate' level of privacy protection. This restriction could potentially hamper the actions of multi-national organizations operating both within the European Union and in non-E.U. nations such as the USA or Japan. This has led to a variety of responses from other governments. Canada and Japan (as well as others) have adopted laws with similar provisions to the Data Protection Directive. The U.S. however, resisted introducing such a comprehensive law and instead proposed a compromise (the Safe Harbor program) [13], which allows U.S. organizations to certify themselves as providing 'adequate' protection.

In an effort to determine the effectiveness of comprehensive privacy legislation, Bowie and Jamal [3] have recently undertaken a comparison of the privacy practices of organizations in the U.S. and U.K. Their study observed the extent that websites' human-readable privacy statements disclose their usage of cookies and analyzed whether the websites *actually* respected requests to opt-out of e-mail marketing. Their study showed that for both issues that U.S. websites "outperformed" their U.K. counterparts. Markel [4] investigates whether the stated actions for those organizations certified as Safe Harbor compliant satisfied the principles set forth by the program. Markel found that 19 out of the 20 randomly selected organizations failed to satisfy the Safe Harbor requirements. A CIPPIC report [2] surveying 72 Canadian websites found widespread non-compliance with PIPEDA. For instance, only a third of surveyed organizations stated they do not share information with third parties for purposes beyond the completion of current activities. Additionally, 35% of the surveyed organizations did not respond to information access requests.

These findings suggest that many of these legal privacy protections have been relatively ineffective at protecting online privacy. However, these surveys are limited in their scope and sample size. Our contribution in this paper is a much broader study of the effects of privacy legislation on websites within a jurisdiction, and a comparison across different jurisdictions with different approaches to privacy protection. The scope of our survey, as well as the sample size, makes this the most comprehensive study of its kind that we are aware of. We have examined more than 3,000 different websites, whereas previous surveys have examined fewer than 100 websites. Furthermore, our analysis covers websites from 13 nations; these include E.U. nations, non-E.U. nations with comprehensive privacy laws, nations without comprehensive privacy laws, and the USA (whose Safe Harbor program is unique). Hence, we believe that this work represents the most comprehensive survey of the question to-date; and that the presented population is sufficiently large and diverse to allow statistical valid conclusions to be drawn subject to the normal limitations presented by any survey without a completely defined sampling framework. The point is discussed further in later sections.

4.2 P3P

P3P [5] became an official recommendation of the W3C in 2002 and is the Internet's official standard for the automated handling of privacy policies. P3P was intended to lessen the burden on Internet users who wish to protect their privacy. Before P3P, privacy conscious Internet users were required to locate the Human Readable Privacy Policy (HRPP), read this potentially lengthy document, determine what elements were of concern, and determine what appropriate steps were required to mitigate the perceived risks. P3P was intended to automate this process and to only provide focused information relevant to the user. P3P requires users to have a P3P user agent installed on their computer and for websites to post a P3P policy document. The P3P agent may be embedded in a browser (such as Internet Explorer 6.0), offered as a browser extension (such as AT&T's Privacy Bird) [14], or act as a proxy server that filters Internet requests (such as the JRC P3P Proxy service) [15].

There are two types of P3P policies. Full P3P policies provide an approximate mapping from a HRPP to an XML document, based on an XML schema published by W3C as a part of the P3P 1.0 specification [16], which is an official W3C recommendation. Compact P3P policies, on the other hand, only contain information relevant to cookies set during an HTTP request/response sequence. Cookies are a significant concern for P3P architects since they allow users to be tracked across websites and can be used to store sensitive information. Since multiple cookies from first and third parties can be set during a HTTP transaction, multiple privacy policies may need to be requested to determine how the information collected will be used. To limit bandwidth usage and time delays, P3P compact policies are simply a string transmitted in the HTTP response header. The P3P specification contains a complete description of both full and compact policies [16].The P3P 1.1 standard has been published as a Working Group note; however, the W3C has decided not to proceed with the candidate recommendation process as of November 2006 [5].

A number of surveys have been conducted that examine the state of P3P adoption. An AT&T survey in 2003, which examined over 5,700 websites drawn from various public rankings of popularity as well as web crawls, found an adoption rate of approximately 9.4% [17]. The most recent surveys found P3P adoption to be relatively low. Egleman *et al.* [18] found that of 80,427 unique websites, 3,846 posted privacy policies in October 2005. In a separate survey of nearly 100,000 websites, Reay *et al.* [19] found that overall P3P adoption was around 3.5% in November 2005. (Website popularity appeared to be related to an increased adoption of P3P; the top 1000 sites had an adoption rate of ~15%.) While P3P is certainly not universally adopted, harvesting P3P policies is the only feasible method available to undertake a broad survey of website privacy policies. A manual analysis of HRPPs posted on the Internet's most popular 100,000 websites using the method employed by Earp *et al.* [20] would require approximately 800 person years to complete!

4.3 Survey Methodology

In order to test the adherence of websites to national privacy legislation, a suitable population of websites must first be identified. Such a population of websites must possess the following characteristics:

- Be large enough to ensure that a sufficient number of P3P policies can be retrieved even if P3P adoption is relatively low;
- Include websites from a broad range of nations;
- Consist of popular websites to ensure that the sample is representative of what average Internet users encounter [17].

In October 2005, Alexa [21], a subsidiary of Amazon.com, provided the authors with a copy of their Top 100,000 list. The list is composed of the most popular 100,000 websites on the Internet as ranked by Alexa.com, based upon the geometric mean of the number of individuals visiting a site and the number of pages they access while on the site. According to Alexa.com, websites which do not belong to this list have less than a 0.00125% change of being visited by the average Internet user [21].

The geographic location of these websites was derived by determining their IP address using the Linux 'host' program; then comparing this address with a database of addresses purchased from IP2Location [22] which maps IP addresses to a particular nation. This approach is required since the country code top level domains are not reliable indicators of the actual location that websites are hosted from [23]. IP2Location claim that their accuracy is above 95% [22]. Through this method, it was determine that the Top 100,000 list contained websites from 130 different nations, and over 3,000 websites posted P3P policies. Thus, it is believed that the list satisfies all of the requirements for the survey outlined above. Note that we are using the physical location of the server to determine the legal jurisdiction that website belongs to. This is a debatable point under international law; however, no better surrogate for determining the applicable jurisdiction is available.

A limitation of this approach is that contextual information pertaining to transactions involving sensitive information is limited. For instance, many privacy laws contain provisions which allow for the processing and distribution of sensitive pieces of information under rare circumstances, such as when a person's life is at stake. P3P cannot describe such extraordinary circumstances [24].

4.3.1 Data Collection

P3P policies were collected and analyzed for validity using the official P3P validator tool provided by the World Wide Web Consortium (W3C). The P3P validator tool retrieves both full and compact policies, and then tests for adherence to the P3P XML Schema. P3P adoption statistics will be provided for all retrieved policies, but only valid policies will be analyzed for adherence to the privacy principles governing populations of websites. It is important to note that the Internet contains a significant number of sites where the P3P document contains errors [19], and that these sites are excluded as their XML P3P documents cannot be reliably parsed.

4.3.2 Data Analysis

Before undertaking the survey, requirements for data analysis were developed to ensure rigor:

- P3P policies must be compared using a standardized approach;
- The standardized approach should be able to map privacy principles onto P3P

policies;

We propose to meet these requirements through the application of well defined rule sets and rigorous statistical analysis. The rule sets must be tailored to individual nations since privacy principles vary between jurisdictions. Further, since privacy concerns vary widely between cultures and the semantic encoding of legal documents is a non-trivial problem, it is unreasonable to expect a standardized language such as P3P to be able to comprehensively express all information a HRPP may contain. Due to these limitations, the rule sets, and hence our analysis, will be limited to those legal mandates that P3P can describe precisely. It should be noted that our analysis is concerned with the letter of the law and hence, does not consider recent case law which may subtly modify the interpretation of specific elements. However, since our analysis pertains to rules built upon legal requirements with little conceptual uncertainty, this problem should be minimal.

4.4 Rule Set Development

Since full and compact policies differ substantially in structure, the rule sets will be defined using the ABNF [25] notation, which was used in the definition of the P3P protocol [16]. To test a P3P policy, the rule sets must be mapped onto an appropriate technology. 'A P3P Preference Exchange Language' (APPEL) [26], which is a W3C working draft, is the official method for expressing rule sets that discriminate between acceptable and unacceptable policies. However, APPEL has come under significant criticism for interactions between APPEL and P3P, which may lead to policies being interpreted incorrectly [27, 28]. Agrawal *et al.* [27] have proposed that XPref can overcome these limitations; this is the technological approach that will be applied for the analysis of full P3P policies. The analysis of P3P compact policies will use regular expressions [29] to identify relevant information elements within the HTTP header string.

4.4.1 Preliminary Results

Of the 100,000 websites in the Alexa 100,000 list, 94,941 were available in November 2006, which is similar to the availability in [19]. In total, 3,282 full and 3,089 compact P3P policies were retrieved. This amounts to an adoption rate of 3.46% for full policies and 3.25% for compact policies, which is similar to the adoption rates observed in [19]. Additionally, a large number of full and compact policies contain errors similar to the rates observed in [19]; 995 (30.3%) full and 885 (28.7%) compact policies contained errors either in their basic XML syntax, or are non-compliant with the P3P 1.0 Schema. To ensure rigor, only valid policies will be analyzed (analysis of syntactically invalid policies would be methodologically questionable). This leaves 2287 full policies and 2204 compact policies in our sample; which equates to maximum margins of error of 1.93% and 1.97% respectively, at the standard 95% limit. These maximal error statements compare very favorably with many results from surveys published within the academic literature. A coverage bias may exist in our dataset, but we have no way to detect or quantify such. Certainly there is a self-selection bias in our data; however, increased sample sizes do not eliminate such a bias (for example, the same self-selection bias would be present in the 83% of websites that post human-readable privacy policies [30]). In any event, the effect of this bias should be to increase the apparent frequency of compliance with legal mandates, as the sites we study have already invested time & effort in crafting a P3P policy. Given our compliance results, we feel that the self-selection bias is not a major threat to validity; and unfortunately no mechanism exists to control it. Of the 131 nations represented in the Top 100,000 list, 17 nations host 95% of the contactable websites from within the list. Websites from these nations also post the vast majority of P3P policies. To ensure adequate sample sizes, a Pareto analysis with a cutoff point of 95% was used to select those nations with the greatest number of full policies (Table 4-1).

We also determined that only 224 of the 777 websites currently certified as Safe Harbor compliant belonged to the Alexa 100,000 list. Hence, we elected to create a second U.S. population of websites from the Safe Harbor list. In total, 745 of the 777 certified sites were available. Of these 745 contactable websites, 26 posted P3P policies. Additionally, of the 224 websites in the overlap, 219 were available and 19 posted P3P policies.

4.4.2 P3P Rule Sets for Legal Obligations

The populations of websites represented in Table 4-1 originate from a range of nations including some which have not adopted comprehensive privacy legislation. Rule sets will only be derived for nations with comprehensive privacy legislation or the Safe Harbor program.

Table 4-1: P3P Adoption

	Total Available Sites	Total valid Full P3P policies		Total valid Compact P3P policies		Valid Full and Compact policies	
U.S.	43767	1525	3.48%	1437	3.28%	415	0.95%
Japan	6569	55	0.84%	53	0.81%	15	0.23%
U.K	3442	168	4.88%	124	3.54%	47	1.37%
Canada	2478	97	3.91%	66	2.66%	26	1.05%
Germany	2352	80	3.40%	44	1.87%	17	0.72%
France	1817	38	2.09%	47	2.59%	13	0.72%
S. Korea	1675	13	0.78%	105	6.27%	2	0.12%
Holland	1489	65	4.37%	76	5.10%	36	2.42%
Spain	1201	16	1.33%	22	1.83%	4	0.33%
Australia	746	27	3.62%	23	3.08%	8	1.07%
Russia	742	25	3.37%	37	4.99%	20	2.70%
Sweden	691	16	2.32%	16	2.32%	2	0.29%
Demark	419	19	4.53%	14	3.34%	6	1.43%

4.4.2.1 European Union

The European Union member nations that will be analyzed are the United Kingdom, Germany, France, Netherlands, Spain, Sweden, and Demark. All of these nations have either implemented or modified existing legislation to meet the requirements set forth by the Data Protection Directive [11]. It should be noted that the Data Protection Directive only sets a base level of protection. Individual nations may implement legislation that exceeds the base protections required by the Directive. Our analysis only includes requirements explicitly set forth in the Directive, thereby ensuring that it remains conservative.

The Joint Resource Center (JRC), which is the science and technology reference service for the European Commission, has developed an APPEL rule set which tests P3P policies for adherence to the EU Data Directive. This rule set is contained within the JRC Policy Editing Workbench [15]. This rule set, however, provides a very strict definition of the Data Protection Directive, since it is designed to ensure that P3P policies comply with the EU directive. There are instances when a policy could fail the test proposed by the JRC rule set and still comply with the Directive. For instance, the rule set developed by the JRC states that websites can only retain information to complete a well defined stated purpose. However, there may be instances where information must be retained for a period mandated by law. For example, globalpositioningsystems.co.uk, a British website, states in their P3P policy that they only retain information for a legally mandated period of time. Under the JRC rule set, this P3P policy would be rejected because they did not state they retain information to complete a stated purpose. Thus, while the JRC rule set provides the foundation for Rules 1-5, minor modifications have been made to the rules to ensure the analysis remains conservative. The five derived rules are:

- Rule 1 (Figure 4-1) tests adherence to Article 14.b of the Directive, which provides individuals with the right to object to having their information used for direct marketing purposes. Information can be used for direct marketing purposes when the 'contact' or 'telemarketing' tags are asserted. When the attribute 'always,' or no attribute, is present in conjunction with a 'contact' or 'telemarketing' tag, users will not be provided with the option to opt-out or opt-in of the data collection.
- Rule 2 (Figure 4-2) tests adherence to Article 12, which provides individuals with the right to access all personal information stored by an organization. The 'all' tag must be asserted in conjunction with the 'access' tag.
- Rule 3 (Figure 4-3) tests adherence to Article 6.1.b, which states that personal data must be collected for specified, explicit, and legitimate purposes. This requires assertion of one or more of the 'stated-purpose', 'business-practices', 'legal-requirement', or 'no-retention' tags.
- Rule 4 (Figure 4-4) tests adherence to Articles 10.a and 11.a, which requires the disclosure of the identity of the data controller, this could include the state, province, street, postal code, or country of the organization.
- Rule 5 (Figure 4-5) tests adherence to Article 25, which allows the exportation of data to an organization in another country only if the other country ensures an adequate level of protection. Assertion of the 'ours' or 'same' tags will pass this rule. The 'delivery' tag poses special problems; contact information must often be passed to a delivery company to complete a transaction. However, the 'delivery' tag does not mean that the delivery company follows the same privacy practices as the vendor; in fact, it is by definition silent on this point. We choose to relax Rule 5 for this case,

and permit assertion of the 'delivery' tag to pass, as delivery is an essential service for any e-commerce company.

Rule 1 = <PURPOSE>

*1*purposevalue*

</PURPOSE>

purposevalue = <contact [required]/> | <telemarketing [required]/>

required = 'always' | ' '

Figure 4-1: Rule 1

Rule 2 = <ACCESS>

*1*1access_allowed*

</ACCESS>

access_allowed = <all/>

Figure 4-2: Rule 2

Rule 3 = <RETENTION>

*1*1retention_period*

</RETENTION>

retention_period = <no-retention/> | <stated-purpose/> | <legal-requirement/> | <business-practice/>

Figure 4-3: Rule 3

```
Rule 4 = <POLICY>
<ENTITY>
<DATA-GROUP>
1*data |
</DATA-GROUP>
<ENTITY/>
</POLICY>
data = <DATA [reference]/>
reference = 'business.contact-info.postal.street' | 'business.contactinfo.postal.city' |
business.contact-info.stateprov' | business.contact-info.postal.postalcode' | business.contact-
info.country'
```

Figure 4-4: Rule 4

```
Rule 5 = <RECIPIENT>
1*recipient
</RECIPIENT>
recipient = <ours/> | <delivery/> | <same/>
```

Figure 4-5: Rule 5

4.4.2.2 Canada

Canada's national privacy law PIPEDA [12], is considered by the European Union to offer 'adequate' protection. Canada's law does however differ on two key points. First, Clause 4.5.2 only recommends that organizations develop guidelines for minimum and maximum retention periods. Thus, when a lack of adherence to Rule 3 is identified, we can only be state that websites are not adhering to recommended practice. Second, Clause 4.1.3 of PIPEDA states that:

> "An organization is responsible for personal information in its possession or custody, including information that has been transferred to a third party for processing. The organization shall use contractual or other means to provide a comparable level of protection while the information is being processed by a third party."

PIPPEDA thus requires a data controller to ensure that *any* third party (not just those in other countries) provides "comparable" privacy protections to the data controller themselves. P3P, however, cannot capture this subtle difference, and so we make no changes for Canadian laws; this may lead to our overstating Canadian compliance with Rule 5, but no other defensible approach exists.

4.4.2.3 Australia

The Privacy Act [31] is Australia's comprehensive privacy law, which offers many of the same protections afforded by the Data Protection Directive. For instance, Clause 2.1 of Schedule 3 allows organizations to use information for telemarketing as long as they provide an opt-out opportunity (Rule 1). Additionally, Clause 6.1 of Schedule 3 states that people should generally be permitted access to information about themselves (Rule 2). Clause 4.2 of Schedule 3 states that organizations must take reasonable steps to destroy or anonymize personal information if it is no longer required (Rule 3). Clause 1 of Schedule 3 also requires organizations to identify themselves and provide contact information (Rule 4). However, while the Privacy Act offers many of the same protections as the EU data directive, the EU Commission has not yet affirmed that the

Australian Privacy law provides 'adequate' protections. Again, we add no further rules for Australian law, due to its similarity to the Data Protection Directive.

4.4.2.4 Japan

To date Japan's comprehensive privacy legislation has also not been affirmed as 'adequate' by the European Commission. Japan's privacy legislation is composed of one central piece of legislation the 'Act Concerning the Protection of Personal Information' followed by four supporting laws [32]. We will not develop tests pertaining to Japanese legislation, as we do not currently have access to an official English translation of this legislation.

4.4.2.5 United States

Websites originating from the United States fall into two categories: those which are self certified as being compliant with the Safe Harbor program [13] and those that are not. Websites certified as compliant with the Safe Harbor principles should satisfy the requirements set forth in the Data Protection Directive. However, some differences do exist. For instance, the "choice principle" states that individuals should be provided with the option to opt-out of any information disclosures to a third party. However, when the information is sensitive such as ethnic origin, political, or religious beliefs, individuals should be presented with an explicit opt-in option. These two requirements lead to the development of Rules 6 (Figure 4-6) and 7 (Figure 4-7). Rule 6 specifies that the recipient of the collected information may be any third party as long as opt-in or opt-out consent is provided. Rule 7 specifies that if 'health', 'political', or 'demographic' data is being collected, then the information can only be shared with other organizations if the user opts into sharing the information.

```
Rule 6 = <RECIPIENT>
*recipient
</RECIPIENT>
recipient = <delivery [required]/> | <same [required]/> | <other-recipient [required]/> |
<unrelated [required]/> | <public [required]/>
required = 'opt-out' | 'opt-in'
```

Figure 4-6: Rule 6

```
Rule 7 =<STATEMENT>
          <DATA-GROUP><DATA>
          <CATEGORIES>category</CATEGORIES>
        </DATA> </DATA-GROUP>
<RECIPIENT>recipient</RECIPIENT>
</STATEMENT>
category = <health/> | <political/> | <demographic/>
recipient = <ours/> | <delivery [required]/> | <same [required]/> | <other-recipient
[required]/> | <unrelated [required]/> | <public [required]/>
required =  'opt-in '
```

Figure 4-7: Rule 7

4.4.2.6 Russia and Korea

Since neither Russia nor Korea implement comprehensive privacy legislation [33-35], websites from these nations will be treated as control groups, as will the US websites which are not members of the Safe Harbor program.

4.5 Data Analysis and Results

Table 4-1 describes the number of European Union websites (as given in Table 4-1) which pass Rules 1-7. Since our dataset is comprised of statements from both full and compact policies, a three way analysis is undertaken for each nation (websites are analyzed depending upon whether they contain a full policy, a compact policy, or both policies). Rule 4 is not applied when only compact policies are examined since compact policies cannot express the relevant information. Since it would appear reasonable to expect low frequency values for various rules (most websites should pass the various rules), we chose to apply Fisher's Exact test instead of the Chi-Square test with Yates correction [36]. This choice ensures our analysis is robust against the occurrences of low-frequency values, which the Chi-Square test has difficulty accommodating [37]. When two populations are being tested against each other, effect size tests will be employed to ensure that statistically significant differences exhibit a medium or large operational effect. Odds-ratio effect sizes [38] will be calculated and then converted into a Cohen's d [38, 39] effect size through the method proposed by Hasselblad and Hedges [40]. The interpretation of the Cohen's d effect sizes will follow the advice provided by Cohen ($d \geq 0.2$ indicates a small effect, $d \geq 0.5$ is a medium effect, and $d \geq 0.8$ indicate a large effect) [38, 39]. Since the Odds-ratio effect size is ill-defined when contingency tables contain cells with an observed frequency of zero, 0.5 will be added to all cells of the contingency tables in accordance with standard practice [41].

Table 4-2: Adherence Results for E.U. Nations

		# of websites	Rule 1		Rule 2		Rule 3		Rule 4		Rule 5		Rule 6		Rule 7		Pass Rules 1-5		Pass Rules 1-7	
			Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)
Both	Denmark	6	100.0	0.0	0.0	100.0	33.3	66.7	83.3	16.7	83.3	16.7	83.3	16.7	83.3	16.7	0.0	100.0	0.0	100.0
	France	13	100.0	0.0	23.1	76.9	15.4	84.6	92.3	7.7	53.8	46.2	46.2	53.8	53.8	46.2	7.7	92.3	7.7	92.3
	Germany	17	94.1	5.9	0.0	100.0	41.2	58.8	100.0	0.0	82.4	17.6	94.1	5.9	100.0	0.0	0.0	100.0	0.0	100.0
	Holland	36	97.2	2.8	2.8	97.2	36.1	63.9	80.6	19.4	88.9	11.1	94.4	5.6	97.2	2.8	2.8	97.2	2.8	97.2
	Spain	4	75.0	25.0	50.0	50.0	0.0	100.0	75.0	25.0	100.0	0.0	50.0	50.0	50.0	50.0	0.0	100.0	0.0	100.0
	Sweden	2	100.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	50.0	50.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0
	U.K.	49	87.8	12.2	16.3	83.7	42.9	57.1	93.9	6.1	71.4	28.6	67.3	32.7	77.6	22.4	12.2	87.8	10.2	89.8
	p-value		0.353		0.022		0.183		0.163		0.101		0.001		0.001		0.512		0.639	
Full	Denmark	19	84.2	15.8	5.3	94.7	36.8	63.2	89.5	10.5	84.2	15.8	84.2	15.8	89.5	10.5	5.3	94.7	5.3	94.7
	France	38	100.0	0.0	15.8	84.2	42.1	57.9	92.1	7.9	97.4	2.6	89.5	10.5	100.0	0.0	2.6	97.4	2.6	97.4
	Germany	80	93.8	6.3	3.8	96.3	53.8	46.3	92.5	7.5	86.3	13.8	93.8	6.3	97.5	2.5	2.5	97.5	2.5	97.5
	Holland	66	81.8	16.7	19.7	78.8	28.8	69.7	86.4	12.1	90.9	7.6	81.8	16.7	95.5	3.0	1.5	97.0	1.5	97.0
	Spain	16	93.8	6.3	18.8	81.3	25.0	75.0	93.8	6.3	93.8	6.3	87.5	12.5	87.5	12.5	0.0	100.0	0.0	100.0
	Sweden	16	68.8	31.3	6.3	93.8	37.5	62.5	93.8	6.3	93.8	6.3	93.8	6.3	93.8	6.3	0.0	100.0	0.0	100.0
	U.K.	170	84.1	15.9	24.1	75.9	45.9	54.1	95.3	4.7	80.0	20.0	65.9	34.1	81.8	18.2	8.2	91.8	5.3	94.7
	p-value		0.004		0.001		0.061		0.467		0.040		0.000		0.000		0.285		0.793	
Compact	Denmark	14	92.9	7.1	7.1	92.9	42.9	57.1	-	-	85.7	14.3	78.6	21.4	85.7	14.3	7.1	92.9	7.1	92.9
	France	47	100.0	0.0	12.8	87.2	44.7	55.3	-	-	70.2	29.8	85.1	14.9	76.6	23.4	12.8	87.2	12.8	87.2
	Germany	44	100.0	0.0	20.5	79.5	50.0	50.0	-	-	84.1	15.9	95.5	4.5	93.2	6.8	18.2	81.8	18.2	81.8
	Holland	76	100.0	0.0	5.3	94.7	44.7	55.3	-	-	94.7	5.3	96.1	3.9	96.1	3.9	5.3	94.7	5.3	94.7
	Spain	22	100.0	0.0	22.7	77.3	59.1	40.9	-	-	100.0	0.0	86.4	13.6	90.9	9.1	13.6	86.4	9.1	90.9
	Sweden	16	100.0	0.0	31.3	68.8	56.3	43.8	-	-	81.3	18.8	93.8	6.3	87.5	12.5	25.0	75.0	25.0	75.0
	U.K.	124	98.4	1.6	16.1	83.9	43.5	56.5	-	-	87.9	12.1	88.7	11.3	89.5	10.5	4.8	95.2	4.0	96.0
	p-value		0.288		0.036		0.825				0.003		0.126		0.044		0.018		0.011	
All E.U Nations	Both	140	92.1	7.9	10.0	90.0	34.3	65.7	89.3	10.7	80.0	20.0	79.3	20.7	85.0	15.0	5.7	94.3	5.0	95.0
	Full	452	87.2	12.8	16.6	83.4	41.2	58.8	92.7	7.3	87.4	12.6	79.4	20.6	91.2	8.8	4.4	95.6	3.3	96.7
	Compact	383	98.4	1.6	13.3	86.7	45.7	54.3	-	-	87.7	12.3	90.9	9.1	89.8	10.2	8.4	91.6	7.8	92.2

Note: The backgrounds of mandatory rules are shaded

The results of Table 4-2 show a number of inconsistencies. First, while adherence to Rule 1 for websites with full, compact and both types of policies is generally 'good', adherence for Swedish websites with full policies appears problematic. This result was unexpected since the Swedish Personal Data Act explicitly states that users must be granted the option of at least opting out of having their personal information used for direct marketing [42]. Given that it is relatively easy to frame questions to ensure that individuals will not opt-out [43], it is surprising that these websites declare privacy policies that contradict the law.

Table 4-2 also indicates that adherence to Rule 2 is generally 'poor'. No nation with more than 10 websites in any of the three groups exhibits an adherence rate above 50% for all three groupings. Furthermore, statistically significant differences in adherence between nations was identified in all 3 groupings of websites. Additionally, the nation with the highest adherence rate varied for each grouping of websites: Spain (Both), U.K. (Full), and Sweden (Compact). We are currently unable to explain why such a chaotic environment would exist for access privileges.

The results of Table 4-2 indicate that adherence to Rule 3 is also generally 'poor'. However, German websites providing full policies appear to break with this trend and appear far more likely to pass Rule 3 than their E.U. counterparts. While a statistically significant difference in the actions of all nations belonging to the group was not identified, we investigate whether German websites specifically differ from any of their European peers in Table 4-3. Those results indicate that a statistically significant difference ($p < 0.004$) with a medium operational effect ($d = 0.561$) exists between German and Dutch websites. This relatively high rate of adherence for German websites

Table 4-3: Comparison of German websites with other E.U. Nations for Rule 3

		Both		Full		Compact	
		pass	fail	pass	fail	pass	fail
German	count	7	10	43	37	22	22
Denmark	count	2	4	7	12	6	8
	p	1.000		0.211		0.762	
	d	0.139		0.363		0.148	
France	count	2	11	16	22	21	26
	p	0.229		0.325		0.677	
	d	0.656		0.253		0.115	
Holland	count	13	23	19	46	34	42
	p	0.768		0.004		0.704	
	d	0.120		0.561		0.115	
Spain	count	0	4	4	12	13	9
	p	0.255		0.054		0.603	
	d	1.026		0.645		0.194	
Sweden	count	2	0	6	10	9	7
	p	0.211		0.281		0.774	
	d	1.073		0.346		0.130	
U.K	count	21	28	78	92	54	70
	p	1.000		0.279		0.485	
	d	0.030		0.172		0.142	

NOTE: The background of statistically significant differences is shaded

may be the result of the strict limitations set forth by the Federal Constitutional Court regarding the retention of information for purposes beyond concluding contact and legal requirements [44].

The results of Table 4-2 also show that while adherence rates to Rules 4, 5, 6 and 7 are generally ‘good’, many statistically significant differences exist, suggesting a large degree of heterogeneity between the nations under analysis. Furthermore, adherence rates of French websites that provide both full and compact policies to Rules 5, 6, and 7 are ‘questionable’. We are surprised by this finding since France institutes a licensing bureau [45] that requires all organizations wishing to undertake data collection and analysis to register their actions.

Table 4-2 also indicates that the number of websites passing Rules 1-5 (or Rules 1-7) simultaneously is ‘extremely low’. Through a visual inspection of Table 4-2, it appears that British websites have a higher rate of adherence than their European counterparts for websites posting full and both types of policies. We investigate this possibility in Table 4-4. These results indicate that no statistically significant differences exist between British websites and their E.U. counterparts who adopt full and both types of policies. There are however a number of medium to large operational effects. It is interesting to note that this apparent increase in adherence does not transfer to the group of U.K. websites posting compact policies; statistically significant differences with large operational effects were found between these sites and German and Swedish websites. We are currently unable to explain why such a difference would exist between websites posting full policies and those posting compact policies in the U.K.

These findings indicate that a significant degree of heterogeneity exists within the stated actions of websites within the E.U. The ‘poor’ adherence to Rules 2 and 3 is particularly concerning. It could be possible that the lack of adherence to Rules 2 and 3 are the result of their potentially large influence on the business practices of an organization. It would appear reasonable to expect that organizations would want to retain information for as long as possible so that it may be used for future marketing campaigns and customer relationship management (Rule 3). However, such actions can significantly affect the privacy of data subjects if that information is stolen or used for purposes beyond that the data subject authorized. It would also appear reasonable to expect that organizations would want to limit the access provided to collected information since additional customer service representatives would need to be employed to handle the access to

Table 4-4: Comparison of the U.K. With Other E.U. Websites

		Both				Full				Compact			
		Rules 1-5		Rules 1-7		Rules 1-5		Rules 1-7		Rules 1-3 and 5		Rules 1-3 and 5-7	
		pass	fail	pass	fail	pass	fail	pass	fail	pass	fail	Pass	fail
U.K.	count	6	43	5	44	14	156	9	161	6	118	5	119
Denmark	count	0	6	0	6	1	18	1	18	1	13	1	13
	p	1.000		1.000		1.000		1.000		0.535		0.480	
	d	0.366		0.261		0.074		0.177		0.389		0.486	
France	count	1	12	1	12	1	37	1	37	6	41	6	41
	p	1.000		1.000		0.315		0.693		0.093		0.073	
	d	0.121		0.016		0.463		0.213		0.578		0.675	
Germany	count	0	17	0	17	2	78	2	78	8	36	8	36
	p	0.326		0.317		0.101		0.511		0.010		0.006	
	d	0.912		0.807		0.589		0.338		0.797		0.894	
Holland	count	1	35	1	35	1	64	1	64	4	72	4	72
	p	0.230		0.236		0.074		0.292		1.000		0.733	
	d	0.696		0.592		0.762		0.512		0.068		0.165	
Spain	count	0	4	0	4	0	16	0	16	3	19	2	20
	p	1.000		1.000		0.614		1.000		0.136		0.284	
	d	0.163		0.059		0.616		0.366		0.654		0.537	
Sweden	count	0	2	0	2	0	16	0	16	4	12	4	12
	p	1.000		1.000		0.614		1.000		0.016		0.010	
	d	0.161		0.265		0.616		0.366		1.037		1.134	

NOTE: The background of statistically significant differences is shaded

information requests. We are however limited in our ability to explore these possibilities, as such motivations are not publicly posted on the Web.

To provide further insight into any significant deviations from the European norm, an amalgamated E.U. population consisting of websites from all 25 E.U. nations is presented in the last line of Table 4-2. Tables 4-5 to 4-7 display the results when this amalgamated population is compared against individual nations for websites posting both full and compact policies (Table 4-5), full policies (Table 4-6), and compact policies (Table 4-7).

Table 4-5: Differences between the Amalgamated E.U. Population and Individual E.U. Nations using Information from both Full and Compact Policies

		Rule 1	Rule 2	Rule 3	Rule 4	Rule 5	Rule 6	Rule 7	Rules 1..5	Rules 1..7
Denmark	*p*	1.000	1.000	1.000	0.508	1.000	1.000	1.000	1.000	1.000
	d	0.079	0.220	0.032	0.437	0.041	0.017	0.229	0.100	0.173
France	*p*	0.600	0.161	0.223	1.000	0.041	0.013	0.013	0.560	0.517
	d	0.482	0.589	0.485	0.016	0.678	0.811	0.866	0.345	0.418
Germany	*p*	1.000	0.367	0.597	0.374	1.000	0.198	0.131	0.600	1.000
	d	0.013	0.766	0.170	0.807	0.027	0.589	1.014	0.446	0.373
Holland	*p*	0.464	0.312	0.846	0.165	0.332	0.047	0.050	0.688	1.000
	d	0.409	0.550	0.050	0.398	0.333	0.713	0.798	0.230	0.157
Spain	*p*	0.297	0.061	0.301	0.379	1.000	0.203	0.120	1.000	1.000
	d	0.868	1.194	0.855	0.686	0.454	0.733	0.946	0.303	0.376
Sweden	*p*	1.000	1.000	0.122	1.000	0.368	1.000	0.275	1.000	1.000
	d	0.448	0.307	1.243	0.266	0.757	0.154	0.058	0.627	0.700
U.K.	*p*	0.388	0.299	0.304	0.413	0.234	0.118	0.269	0.200	0.304
	d	0.287	0.320	0.201	0.273	0.263	0.343	0.279	0.466	0.435

NOTE: The background of statistically significant differences is shaded

The results of this analysis suggest a number of large statistically significant deviations exist between the amalgamated European population and individual E.U. nations. For instance statistically significant differences with a medium/large operational effect were identified for French, German, U.K., and Dutch websites for the various groupings of websites (Both, Full, Compact). It is surprising that German websites providing full

Table 4-6: Differences between the Amalgamated E.U. Population and Individual E.U. Nations using Information from Full Policies

		Rule 1	Rule 2	Rule 3	Rule 4	Rule 5	Rule 6	Rule 7	Rules 1..5	Rules 1..7
Denmark	*p*	0.725	0.335	0.814	0.644	0.722	0.776	0.683	0.587	0.488
	d	0.201	0.495	0.084	0.328	0.212	0.110	0.213	0.309	0.474
France	*p*	0.015	1.000	0.865	0.511	0.070	0.201	0.061	1.000	1.000
	d	1.339	0.003	0.026	0.124	0.707	0.378	1.109	0.081	0.084
Germany	*p*	0.131	0.002	0.038	0.819	0.585	0.002	0.069	0.555	0.730
	d	0.388	0.817	0.279	0.056	0.075	0.699	0.615	0.206	0.041
Holland	*p*	0.336	0.483	0.078	0.214	0.311	0.620	0.145	0.498	0.706
	d	0.199	0.141	0.282	0.347	0.255	0.113	0.498	0.380	0.215
Spain	*p*	0.706	0.738	0.300	0.620	0.706	0.751	0.646	1.000	1.000
	d	0.231	0.146	0.366	0.114	0.220	0.224	0.317	0.234	0.069
Sweden	*p*	0.050	0.488	1.000	0.620	0.706	0.214	1.000	1.000	1.000
	d	0.650	0.397	0.067	0.114	0.220	0.543	0.002	0.234	0.069
U.K.	*p*	0.359	0.038	0.317	0.281	0.022	0.001	0.002	0.075	0.251
	d	0.145	0.263	0.107	0.226	0.309	0.384	0.465	0.383	0.297

NOTE: The background of statistically significant differences is shaded

Table 4-7: Differences between the Amalgamated E.U. Population and Individual E.U. Nations using Information from Compact Policies

		Rule 1	Rule 2	Rule 3	Rule 5	Rule 6	Rule 7	Rules 1..5	Rules 1..7
Denmark	*p*	0.224	1.000	1.000	0.687	0.141	0.646	1.000	1.000
	d	1.071	0.179	0.053	0.197	0.610	0.313	0.109	0.148
France	*p*	0.503	1.000	1.000	0.003	0.199	0.014	0.285	0.262
	d	0.228	0.011	0.020	0.623	0.337	0.564	0.298	0.337
Germany	*p*	0.481	0.249	0.634	0.475	0.406	0.602	0.051	0.043
	d	0.192	0.306	0.095	0.197	0.296	0.163	0.517	0.556
Holland	*p*	0.260	0.053	0.900	0.107	0.172	0.124	0.485	0.631
	d	0.491	0.500	0.020	0.448	0.412	0.478	0.212	0.173
Spain	*p*	1.000	0.208	0.273	0.092	0.449	0.713	0.423	0.689
	d	0.184	0.395	0.289	1.014	0.319	0.040	0.373	0.199
Sweden	*p*	0.219	0.059	0.451	0.436	1.000	0.674	0.046	0.038
	d	0.355	0.626	0.226	0.340	0.021	0.231	0.757	0.796
U.K.	*p*	1.000	0.457	0.756	1.000	0.486	1.000	0.241	0.220
	d	0.137	0.135	0.047	0.007	0.147	0.036	0.280	0.338

NOTE: The background of statistically significant differences is shaded

policies were significantly more likely to fail Rule 2 when Germany is considered to offer some of the strictest privacy protections in the E.U [44]. Similarly, it is also surprising that websites providing compact and both types of policies from France were far more likely to fail Rules 5 and 7, since France has one of the longest traditions of enshrining privacy protections in law. The mixed results from compact policies posted on British websites, which were significantly more likely to pass Rule 2, but more likely to fail Rules 5, 6, and 7, further suggest that significant deviations appear to exist within the U.K. populations. The statistically significant increases in the number of compact policies passing Rules 1-7 from Germany and Sweden does bode well for these two nations; note, however, that this seems contrary to the earlier finding that German websites providing full policies were more likely to fail Rule 2.

Tables 4-8, 4-9, and 4-10 describe the adherence to Rules 1-7 across the remaining jurisdictions under analysis, with the U.S. population split into the general U.S. population and the Safe Harbor population. As was observed in Table 4-2, few websites from the various populations in Tables 4-8 to 4-10 satisfied Rules 2 and 3 with the only exception being South Korea. It is surprising that 63.3% of South Korean websites using compact policies (Table 4-10) satisfied Rule 2 when the jurisdiction with the next highest adherence rate is the Safe Harbor group (23.6%). This relatively high adherence rate may be the result of 'The Act on Promotion of Information and Communications Network Utilization and Data Protection,' which came into effect in 2000 and, while it is not comprehensive, does govern the information and telecommunications industries [35]. Additionally, the large variance in adherence rates for Australian websites from 37.0% in the full policy group to 4.3% in the compact policy group is very surprising. This difference was found to be statistically significant with a large operational effect ($p < 0.0064$, d = 1.412) and is partially the result of inconsistent statements between full and compact policies (Table 4-15).

We are also surprised by the relatively high degree of uniformity that appears to exist for Australian websites that provide full and compact policies. The results of Table 4-8 for Rules 1, 2, 4, 6, and 7 indicate that these websites either all pass or all fail these rules. These results could be the result of a homogeneous population of websites and this possibly will be explored in Table 4-16.

Table 4-8: Adherence Results for Non E.U. Nations with Both Full and Compact Policies

		Rule 1		**Rule 2**		**Rule 3**		**Rule 4**		**Rule 5**		**Rule 6**		**Rule 7**		**Passed Rules 1-5**		**Passed Rules 1-7**	
	# of websites	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Pass (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)
Amalgamated E.U. Population																			
E.U.	140	92.1	7.9	10.0	90.0	34.3	65.7	89.3	10.7	80.0	20.0	79.3	20.7	85.0	15.0	5.7	94.3	5.0	95.0
Nations With Comprehensive Privacy Legislation																			
Australia	8	100.0	0.0	0.0	100.0	37.5	62.5	100.0	0.0	62.5	37.5	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0
Canada	26	80.8	19.2	3.8	96.2	19.2	80.8	80.8	19.2	76.9	23.1	84.6	15.4	88.5	11.5	0.0	100.0	0.0	100.0
Japan	15	86.7	13.3	6.7	93.3	33.3	66.7	93.3	6.7	46.7	53.3	86.7	13.3	80.0	20.0	0.0	100.0	0.0	100.0
p-value	-	0.563		1.000		0.455		0.356		0.149		0.716		0.603		1.000		1.000	
Nations Without Comprehensive Privacy Legislation																			
S. Korea	2	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	100.0	0.0	0.0	100.0	0.0	100.0
Russia	20	100.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0
p-value	-	0.091		0.091		0.091		0.091		0.091		0.091		1.000		1.000		1.000	
U.S. Populations																			
Safe Harbor	8	87.5	12.5	37.5	62.5	50.0	50.0	75.0	25.0	87.5	12.5	75.0	25.0	75.0	25.0	12.5	87.5	12.5	87.5
U.S.	415	88.0	12.0	12.3	87.7	37.3	62.7	94.7	5.3	68.4	31.6	74.0	26.0	82.9	17.1	4.1	95.9	4.1	95.9
p-value	-	1.000		0.069		0.481		0.070		0.444		1.000		0.631		0.2959		0.296	

Table 4-9: Adherence Results for Non E.U. Nations with Full Policies

		Rule 1		Rule 2		Rule 3		Rule 4		Rule 5		Rule 6		Rule 7		Passed Rules 1-5		Passed Rules 1-7	
	# of websites	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Pass (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)
Amalgamated E.U. Population																			
E.U.	452	87.2	12.8	16.6	83.4	41.2	58.8	92.7	7.3	87.4	12.6	79.4	20.6	91.2	8.8	4.4	95.6	3.3	96.7
Nations With Comprehensive Privacy Legislation																			
Australia	27	85.2	14.8	37.0	63.0	44.4	55.6	100.0	0.0	59.3	40.7	70.4	29.6	92.6	7.4	3.7	96.3	0.0	100.0
Canada	97	90.7	9.3	17.5	82.5	45.4	54.6	76.3	23.7	86.6	13.4	78.4	21.6	94.8	5.2	1.0	99.0	1.0	99.0
Japan	55	90.9	9.1	3.6	96.4	41.8	58.2	94.5	5.5	65.5	34.5	85.5	14.5	83.6	16.4	0.0	100.0	0.0	100.0
p-value	-	0.695		0.000		0.935		0.000		0.001		0.264		0.065		0.373		1.000	
Nations Without Comprehensive Privacy Legislation																			
S. Korea	13	92.3	7.7	61.5	38.5	76.9	23.1	53.8	46.2	69.2	30.8	61.5	38.5	100.0	0.0	23.1	76.9	0.0	100.0
Russia	25	100.0	0.0	4.0	96.0	12.0	88.0	96.0	4.0	96.0	4.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0
p-value	-	0.342		0.000		0.000		0.004		0.038		0.003		1.000		0.034		1.000	
U.S. Populations																			
Safe Harbor	26	92.3	7.7	30.8	69.2	57.7	42.3	84.6	15.4	80.8	19.2	76.9	23.1	88.5	11.5	7.7	92.3	7.7	92.3
U.S.	1526	87.8	12.2	17.4	82.6	43.6	56.4	89.6	10.4	79.2	20.8	72.7	27.3	83.2	16.8	5.6	74.7	5.0	95.0
p-value	-	0.761		0.112		0.167		0.343		1.000		0.825		0.604		0.704		0.677	

Table 4-10: Adherence Results for Non E.U. Nations with Compact Policies

		Rule 1		**Rule 2**		**Rule 3**		**Rule 5**		**Rule 6**		**Rule 7**		**Passed Rules 1-5**		**Passed Rules 1-7**	
	# of websites	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Pass (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)	Pass (%)	Fail (%)
Amalgamated E.U. Population																	
E.U.	383	98.4	1.6	13.3	86.7	45.7	54.3	87.7	12.3	90.9	9.1	89.8	10.2	8.4	91.6	7.8	92.2
Nations With Comprehensive Privacy Legislation																	
Australia	23	100.0	0.0	4.3	95.7	47.8	52.2	100.0	0.0	95.7	4.3	95.7	4.3	4.3	95.7	4.3	95.7
Canada	66	93.9	6.1	6.1	93.9	36.4	63.6	86.4	13.6	92.4	7.6	95.5	4.5	3.0	97.0	3.0	97.0
Japan	53	94.3	5.7	5.7	94.3	56.6	43.4	66.0	34.0	83.0	17.0	90.6	9.4	1.9	98.1	1.9	98.1
p-value		0.761		1.000		0.087		0.001		0.214		0.591		0.818		0.818	
Nations Without Comprehensive Privacy Legislation																	
S. Korea	105	98.1	1.9	63.8	36.2	14.3	85.7	98.1	1.9	76.2	23.8	79.0	21.0	0.0	100.0	0.0	100.0
Russia	37	100.0	0.0	2.7	97.3	10.8	89.2	94.6	5.4	83.8	16.2	81.1	18.9	0.0	100.0	0.0	100.0
p-value		1.000		0.000		0.781		0.278		0.488		1.000		1.000		1.000	
U.S. Populations																	
Safe Harbor	21	95.2	4.8	19.0	81.0	33.3	66.7	81.0	19.0	81.0	19.0	85.7	14.3	4.8	95.2	4.8	95.2
U.S.	1439	95.3	4.7	10.1	89.9	34.1	65.9	58.2	41.8	87.4	12.6	91.0	9.0	4.4	95.6	3.7	96.3
p-value		1.000		0.263		1.000		0.043		0.328		0.431		0.618		0.549	

With respect to Canadian websites, only 19.2% of websites with both full and compact policies (Table 4-8), 45.4% of websites with full policies (Table 4-9), and 36.4% of websites posting compact policies (Table 4-10) follow the recommendations set forth in PIPEDA that information should not be retained indefinitely (Rule 3). Canadian websites posting full policies do however have the highest adherence rates for Rule 5 ($p < 0.026$) (see Table 4-9). This finding may suggest that Clause 4.1.3 of PIPEDA has been effective at limiting disclosures to third parties following different practices. A CIPPIC study [2] found that about half of the 64 Canadian online retailers they surveyed do share customer information with third parties. However, this does not measure whether or not those companies are following the same or similar privacy practices, merely that the information is shared for purposes beyond the completion of the current transaction. This study also used university law students to assess human-readable privacy policies and to contact the companies for further information.

It is surprising that few statistically significant differences were found in the stated actions of websites from South Korea and Russia which use compact policies (Table 4-10), since neither nation has enacted comprehensive legislation. The only rule where a statistically significant difference was identified was in Rule 2 ($p < 0.001$) where we theorized that this difference could be due to the adoption of the 'The Act on Promotion of Information and Communications Network Utilization and Data Protection.' Furthermore, a significant degree of homogeneity appears to exist in Russian websites similar to what was previously observed in Australian websites. We will investigate this peculiarity in Table 4-16.

The comparison between the general U.S. and the Safe Harbor populations of websites bring into question the effectiveness of the Safe Harbor Program. The only statistically significant difference between the Safe Harbor population and the general U.S. population was in compact policies for Rule 5. It is surprising that no statistically significant differences were identified between the two populations for Rules 6 and 7. This homogeneity could indicate that the self-certification procedure used in the Safe Harbor program is ineffective. Markel [4] also investigated adherence to Rules 6 and 7 within the Safe Harbor community and found that 4 out of 20 organizations failed Rule 6 and 13 out of 20 failed Rule 7. While a 20% failure rate for Rule 6 is roughly comparable with the results of Tables 4-8, 4-9, and 4-10, our results differ substantially from Markel's on Rule 7. We are however unable to ascertain why this difference exists since Markel's method relied upon a subjective interpretation of HRPP documents, rather than automated analysis of P3P documents.

We investigate whether these populations of websites differ significantly from nations governed by the Data Protection Directive in Tables 4-11, 4-12, and 4-13. These tables test whether the amalgamated E.U. population previously discussed differs significantly from the various Non-E.U. nations.

Table 4-11: Differences between the Amalgamated E.U. Population and non-E.U. Nations using Information from both Full and Compact Policies

		Rule 1	Rule 2	Rule 3	Rule 4	Rule 5	Rule 6	Rule 7	Rules 1-5	Rule 1-7
Australia	*p*	1.000	1.000	1.000	0.584	0.365	0.371	0.602	1.000	1.000
	d	0.227	0.368	0.107	0.409	0.508	0.829	0.616	0.048	0.025
Canada	*p*	0.138	0.470	0.171	0.320	0.792	0.789	0.770	0.359	0.598
	d	0.583	0.368	0.396	0.401	0.124	0.154	0.104	0.675	0.602
Japan	*p*	0.365	1.000	1.000	1.000	0.008	0.737	0.705	1.000	1.000
	d	0.405	0.057	0.001	0.098	0.826	0.197	0.244	0.379	0.306
Korea	*p*	0.162	0.201	1.000	0.213	0.368	0.379	0.275	1.000	1.000
	d	1.335	1.194	0.356	1.153	0.757	0.733	0.058	0.627	0.700
Russia	*p*	0.362	0.219	0.000	0.220	0.026	0.026	0.078	0.597	0.598
	d	0.712	0.853	1.691	0.894	1.290	1.314	1.102	0.533	0.460
Safe Harbor	*p*	0.500	0.049	0.452	0.230	1.000	0.674	0.611	0.402	0.366
	d	0.448	0.945	0.356	0.626	0.130	0.206	0.419	0.627	0.700
US	*p*	0.211	0.545	0.544	0.032	0.009	0.257	0.602	0.479	0.635
	d	0.244	0.115	0.072	0.425	0.332	0.159	0.079	0.209	0.136

NOTE: The background of statistically significant differences is shaded

Table 4-12: Differences between the Amalgamated E.U. Population and non-E.U. Nations using Information from Full Policies

		Rule 1	Rule 2	Rule 3	Rule 4	Rule 5	Rule 6	Rule 7	Rules 1-5	Rule 1-7
Australia	*p*	0.767	0.016	0.841	0.243	0.000	0.329	1.000	1.000	1.000
	d	0.141	0.606	0.078	0.816	0.567	0.285	0.001	0.098	0.368
Canada	*p*	0.395	0.881	0.497	0.000	0.867	0.784	0.308	0.148	0.328
	d	0.178	0.046	0.095	0.757	0.052	0.043	0.277	0.615	0.454
Japan	*p*	0.521	0.009	1.000	0.575	0.000	0.372	0.089	0.150	0.390
	d	0.170	0.802	0.018	0.100	0.718	0.206	0.404	0.915	0.755
Korea	*p*	1.000	0.001	0.019	0.000	0.077	0.160	0.615	0.022	1.000
	d	0.117	1.127	0.802	1.315	0.651	0.503	0.537	1.075	0.024
Russia	*p*	0.058	0.154	0.003	0.696	0.342	0.007	0.253	0.615	1.000
	d	1.115	0.653	0.829	0.146	0.477	1.425	0.888	0.487	0.326
Safe Harbor	*p*	0.759	0.104	0.106	0.132	0.362	0.803	0.720	0.339	0.235
	d	0.206	0.459	0.361	0.506	0.312	0.109	0.230	0.423	0.583
US	*p*	0.745	0.723	0.358	0.057	0.000	0.004	0.000	0.055	0.161
	d	0.035	0.028	0.056	0.209	0.328	0.204	0.397	0.259	0.227

NOTE: The background of statistically significant differences is shaded

Table 4-13: Differences between the Amalgamated E.U. Population and non-E.U. Nations using Information from Compact Policies

		Rule 1	Rule 2	Rule 3	Rule 5	Rule 6	Rule 7	Rules 1-3 and 5	Rule 1-3 and 5-7
Australia	*p*	0.297	0.336	1.000	0.092	0.709	0.715	1.000	1.000
	d	0.117	0.465	0.049	1.043	0.234	0.299	0.180	0.142
Canada	*p*	0.045	0.107	0.181	0.692	0.818	0.174	0.204	0.201
	d	0.789	0.422	0.209	0.087	0.072	0.404	0.479	0.441
Japan	*p*	0.084	0.125	0.144	0.000	0.088	1.000	0.159	0.154
	d	0.768	0.443	0.239	0.720	0.408	0.006	0.647	0.609
Korea	*p*	0.684	0.000	0.000	0.001	0.000	0.005	0.001	0.001
	d	0.187	1.338	0.878	0.973	0.625	0.471	1.638	1.600
Russia	*p*	1.000	0.067	0.000	0.288	0.239	0.161	0.097	0.095
	d	0.141	0.731	1.012	0.383	0.389	0.421	1.068	1.030
Safe Harbor	*p*	0.046	0.508	0.368	0.321	0.133	0.470	1.000	1.000
	d	0.798	0.279	0.268	0.331	0.511	0.276	0.129	0.091
US	*p*	0.005	0.079	0.000	0.000	0.075	0.428	0.004	0.001
	d	0.579	0.173	0.267	0.898	0.192	0.077	0.374	0.444

NOTE: The background of statistically significant differences is shaded

The results of Tables 4-11 to 4-13 indicate that no medium/large differences in overall adherence to Rules 1-5 exist between the amalgamated E.U. population and all non-E.U. nations with the exception of Korea. The statistically significant differences between the E.U. and Korean websites posting full ($p < 0.022$, d = 1.075) and compact ($p < 0.001$, d = 1.638) policies were large. It is interesting that a nation with only sectoral privacy legislation exhibits better adherence to the Data Protection Directive than E.U. nations – at least for full policies. However, the results from Korean websites indicate that a significant dichotomy exists between those sites posting full or compact policies. Websites adopting full policies were significantly more likely to pass Rules 1-5 than E.U sites but sites adopting compact policies were significantly less likely to pass Rules 1-5. We are currently unable to explain why such a divergence exists between Korean websites. The effectiveness of the Directive is further brought into question since Russian websites often show a significant improvement over E.U. websites for Rules 5 and 6, although Rule 6 is strictly not covered by the Directive (Tables 4-11 and 4-12). This result is unexpected since the privacy protections afforded within Russia are questionable at best [33] and information misuse is rampant [34].

Since adherence rates are generally quite low, it would appear reasonable to theorize that websites may not be providing sufficient resources toward the development of their policies. If this is the case, it would appear likely that inconsistencies should be found between the contents of full and compact policies. To investigate this possibility Tables 4-14 and 4-15 analyze the consistency in the stated actions of websites who posted both full and compact policies. For each nation, we compare full and compact policies for every rule, for every website that posts both. We report the number of conflicts and their frequency for E.U. nations in Table 4-14, and for non-E.U. nations in Table 4-15. For example, 3 websites from France state in their full policy that they satisfy Rule 3 while their compact policy indicates that they fail the same rule (Table 4-14). It should however be noted that definitive statements of whether conflicts occur when the full policy fails and the compact policy passes cannot reliably be made since compact policies are only required to contain a subset of the information present in the accompanying full policy. For instance, if a website was to collect information through HTML forms and permanently retain this information, while information collected through cookies was only retained to complete the websites stated actions, the full policy would fail Rule 3 while the compact policy could still potentially pass. For this reason, caution is urged when interpreting the results when the full policy fails and the compact policy passes.

The results of Tables 4-14 and 4-15 indicate that while the total number of websites providing both full and compact policies are low, a large proportion of these websites contradict themselves (full policy passes, compact policy fails). This rate of contradiction ranges from a high of 46.2% in France (Table 4-14) to 0.0% in Russia (Table 4-15). It should be noted that nations with fewer than 10 websites adopting both full and compact policies are excluded from this discussion since their results are deemed unreliable. These rates of contradiction are unexpected since it would appear reasonable to hypothesize that websites from France would be more likely to provide resources toward the development of privacy policies, since the protections afforded by French law are much more stringent than Russian protections. We are currently unable to explain why such a difference exists. It is also surprising that almost 20% of U.S. websites contradict themselves on at least one of the seven rules under analysis (Table 4-15).

These inconsistencies could be the result of a lack of attention, or websites attempting to circumvent the P3P agent installed in Internet Explorer 6.0, which can filter cookies based upon P3P compact policy contents. We are however limited in our ability to further investigate these possibilities, since the organizations do not post their motivations on their websites. This finding should, however, at least raise concerns about the quality of information provided to users.

Table 4-14: Number of E.U. Websites who's stated Actions are Inconsistent

	Full	Compact	Rule 1	Rule 2	Rule 3	Rule 5	Rule 6	Rule 7	Total Number Contradicting
E.U. Nations									
Denmark	Pass	Fail	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
	Fail	Pass	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
France	Pass	Fail	0 0.0%	0 0.0%	0 0.0%	6 46.2%	6 46.2%	6 46.2%	6 46.2%
	Fail	Pass	0 0.0%	0 0.0%	3 23.1%	0 0.0%	1 7.7%	0 0.0%	3 23.1%
Germany	Pass	Fail	0 0.0%	1 5.9%	3 17.7%	2 11.8%	0 0.0%	0 0.0%	5 29.4%
	Fail	Pass	1 5.9%	2 11.8%	1 5.9%	1 5.9%	1 5.9%	0 0.0%	6 35.3%
Holland	Pass	Fail	0 0.0%	1 2.8%	2 5.6%	0 0.0%	0 0.0%	0 0.0%	3 8.3%
	Fail	Pass	1 2.8%	0 0.0%	2 5.6%	3 5.9%	2 5.6%	1 2.8%	5 13.9%
Spain	Pass	Fail	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
	Fail	Pass	1 25.0%	0 0.0%	1 25.0%	0 0.0%	1 25.0%	1 25.0%	1 25.0%
Sweden	Pass	Fail	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
	Fail	Pass	0 0.0%	1 50.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 50.0%
U.K.	Pass	Fail	0 0.0%	2 4.1%	0 0.0%	4 8.2%	0 0.0%	3 6.1%	7 14.3%
	Fail	Pass	4 8.2%	3 6.1%	3 6.1%	8 16.3%	7 14.3%	6 12.2%	16 36.6%

NOTE: The background of rules websites are mandated to adhere to are shaded

Table 4-15: Number of Non-E.U. Websites who's stated Actions are Inconsistent

	Full	Compact	Rule 1	Rule 2	Rule 3	Rule 5	Rule 6	Rule 7	Total Number Contradicting
Non-E.U. Nations With Comprehensive Privacy Legislation									
Australia	Pass	Fail	0 0.0%	3 37.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	3 37.5%
	Fail	Pass	0 0.0%	0 0.0%	4 50.0%	3 37.5%	0 0.0%	0 0.0%	7 87.5%
Canada	Pass	Fail	0 0.0%	3 11.5%	2 7.7%	0 0.00%	0 0.00%	1 3.9%	6 23.1%
	Fail	Pass	3 11.5%	0 0.0%	3 11.5%	3 11.5%	1 3.9%	1 3.9%	9 34.6%
Japan	Pass	Fail	1 6.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 6.7%
	Fail	Pass	0 0.0%	0 0.0%	1 6.7%	0 0.0%	0 0.0%	1 6.7%	2 13.3%
Non-E.U. Nations Without Comprehensive Privacy Legislation									
South Korea	Pass	Fail	0 0.0%	0 0.0%	1 50.0%	0 0.0%	0 0.0%	0 0.0%	1 50.0%
	Fail	Pass	1 50.0%	0 0.0%	0 0.0%	1 50.0%	0 0.0%	0 0.0%	2 100.0 %
Russia	Pass	Fail	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
	Fail	Pass	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
U.S. Websites									
Safe Harbor	Pass	Fail	1 12.5%	0 0.0%	0 0.0%	0 0.0%	1 12.5%	1 12.5%	1 12.5%
	Fail	Pass	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
U.S.	Pass	Fail	7 1.7%	2 45.8%	27 6.5%	22 5.3%	16 3.9%	28 6.8%	82 19.8%
	Fail	Pass	31 7.5%	17 4.1%	37 8.9%	54 13.0%	24 5.8%	16 3.9%	128 30.8%

NOTE: The background of rules websites are mandated to adhere to are shaded

To provide greater insight into whether the inconsistencies observed in Tables 4-2, 4-8, 4-9, and 4-10 are the result of outliers or a generally chaotic environment for privacy protection, we propose to test the consistency of the pass/fail response sequences for the websites of each national population. Due to the potentially large number of variables, standard bivariate approaches such as Pearson or Spearman Rank correlations are not applicable. Instead, we will apply Cronbach's Alpha [46-48], and the Intraclass Correlation (ICC) [37, 49-52] on the dichotomous response variables indicating whether a website passes or fails the various rules. Cronbach's Alpha provides a method for rating the internal consistency of a series of responses (website P3P policies) on various subjects of interest (P3P tags). However, Cronbach's Alpha is not a measure of unidimensionality. A dataset can have a high Cronbach's Alpha and still be multidimensional when clusters of highly intercorrelated items exist and weak correlation exists between these clusters [47]. As a result, Cronbach's Alpha will be utilized as an exploratory test indicating whether a population exhibits general internal consistency. Thus, while a weak result indicates no consistency or agreement, a strong result cannot be interpreted as agreement between websites regarding privacy issues. The lower bound of 0.7, suggested by Nunnally and Bernstien [53] and Garson [47], will be used as an indication of general internal consistency.

Intraclass correlation will be utilized to determine the degree of absolute website agreement on the subjects of interest (P3P tags). Intraclass correlation (ICC) should not be confused with standard correlation techniques; ICC allows for analysis concerning both the degree of association as well as the repeatability of the association when the observed variables are dichotomous in nature. The analysis in the current paper utilizes the method generally referred to as Model 3 [51, 52] using individual measurements; Model 3 is appropriate when all subjects of interest are surveyed [51]. This approach is reasonable since this study surveys all popular P3P adopting websites. Generalizations to all popular websites (whether they adopt P3P or not) would require the usage of Model 2 and must be made at the reader's discretion. However, as a cross check, the ICC results for Model 2 were calculated and no significant differences exist between the results from Models 2 and 3.

Interpretations of ICC results differ from those generated through standard correlation methods; the ICC value cannot be viewed as an r^2 value, since it takes into account consistency *as well as* repeatability of results. We will follow the recommendations of Portney and Watkins [51], who state that an ICC above 0.75 is an indicator of good agreement. In practice, ICC values range from 0.00 to 1.00. However, in certain circumstances, the ICC values can range from $\pm\infty$ [51]. This situation occurs when the data set is homogeneous, meaning a lack of significant variance between subjects of interest (pass/fail sequences for full and compact policies); this situation can be discovered by the application of a one-way ANOVA test [50]. When a dataset is found to be homogenous, the test is considered inconclusive.

A limitation of using the ICC method is the lack of clear consensus regarding the calculation of Type I and Type II errors. For example, little guidance exists regarding the choice of a null hypothesis for an ICC test. The choice of $H_0 = 0$, for instance, provides information of little practical importance. Walter *et al* [54] recommended setting H_0 to the minimally acceptable level of reliability. However, little guidance is given for choosing this point without making the tests overly conservative. The calculation of a minimum detectable effect (MDE) [55] is a potential solution. MDE's describe the smallest effect that can be detected at a given statistical power and significance level, thus guarding against these errors. However, no formulation exists for the calculation of a MDE when using ICC. As a result of these limitations, confidence intervals will be calculated as an indicator of the degree of uncertainty in the results [56]. This does limit our ability to analyze consistency, and thus we will make only very conservative statements in this regard.

Tables 4-16, 4-17, and 4-18 analyze the response patterns derived from full, compact, and both full and compact policies, respectively. The results indicate that no nation except Russia exhibits high consistency (ICC > 0.75) in the pass/fail sequences. Furthermore, high consistency was only observed in the groups of Russian websites that adopted full or both types of policies. This high reliability is likely the result of the high degree of homogeneity in Russian full P3P policies previously observed in [57]. That study focused on whether norms of behavior had developed with respect to privacy protection on the Internet, comparing national culture and the tendency to adopt particular privacy protections (rather than comparing the content of policies against specific legal requirements as in the current paper). Oddly enough, the 20 Russian websites posting full and compact policies exhibited perfect consistency (ICC=1.0), whereas the 37 websites posting compact policies do not exhibit good consistency (ICC=0.703). The results of this analysis also indicate that the uniformity we observed in Australian websites for individual rules is not due to the population being highly homogeneous. Currently we are unable to explain why the Australian population exhibits these unique characteristics. It should be noted that while Cronbach's alpha generally varies from 0 to 1, in exceptional

Table 4-16: Intraclass Correlation Results Testing Reliability of Website Adherence to Rules Using P3P Full Policies

				95% confidence interval		F-Test		
Country	Number of Websites	Cronbach's Alpha	Intraclass Correlation	Lower	Upper	Value	df	*p*
E.U. Nations								
Denmark	19	0.942	0.442	0.222	0.803	17.217	6	0.000
French	38	0.980	0.571	0.345	0.869	49.148	6	0.000
German	80	0.991	0.571	0.351	0.867	114.662	6	0.000
Netherlands	65	0.980	0.443	0.240	0.797	51.234	6	0.000
Spain	16	0.938	0.492	0.256	0.834	16.142	6	0.000
Sweden	16	0.947	0.507	0.271	0.841	18.958	6	0.000
UK	170	0.986	0.273	0.132	0.648	70.397	6	0.000
Non-EU nations with Comprehensive Privacy Legislation								
Australia	27	0.894	0.238	0.094	0.625	9.443	6	0.000
Canada	97	0.981	0.354	0.180	0.730	53.431	6	0.000
Japan	55	0.982	0.452	0.247	0.803	56.213	6	0.000
Nations without Comprehensive Privacy Legislation								
Korea	13	0.458	0.068	-0.025	0.413	1.846	6	0.102
Russia	25	0.993	0.855	0.702	0.967	141.517	6	0.000
U.S. Websites								
Safe Harbor	26	0.870	0.202	0.074	0.581	7.678	6	0.000
US	1525	0.996	0.341	0.176	0.716	276.645	6	0.000

Table 4-17: Intraclass Correlation Results Testing Reliability of Website Adherence to Rules Using P3P Compact Policies

				95% confidence interval		F-Test		
Country	Number of Websites	Cronbach's Alpha	Intraclass Correlation	Lower	Upper	Value	df	*p*
E.U. Nations								
Denmark	14	0.933	0.426	0.189	0.828	14.880	5	0.000
French	47	0.979	0.394	0.192	0.800	47.776	5	0.000
German	44	0.980	0.470	0.246	0.845	50.409	5	0.000
Netherlands	76	0.995	0.681	0.449	0.928	190.854	5	0.000
Spain	22	0.947	0.447	0.215	0.837	18.706	5	0.000
Sweden	16	0.893	0.307	0.114	0.748	9.340	5	0.000
UK	124	0.992	0.480	0.261	0.849	127.352	5	0.000
Non-EU nations with Comprehensive Privacy Legislation								
Australia	23	0.984	0.703	0.462	0.936	62.526	5	0.000
Canada	66	0.990	0.594	0.356	0.899	96.472	5	0.000
Japan	53	0.982	0.439	0.225	0.827	56.972	5	0.000
Nations without Comprehensive Privacy Legislation								
Korea	105	0.991	0.435	0.226	0.824	105.773	5	0.000
Russia	37	0.988	0.703	0.469	0.935	86.587	5	0.000
U.S. Websites								
Safe Harbor	21	0.934	0.383	0.170	0.800	15.162	5	0.000
US	1437	0.998	0.477	0.261	0.846	509.024	5	0.000

Table 4-18: Intraclass Correlation Results Testing Reliability of Website Adherence to Rules Using both Full and Compact Policies

				95% confidence interval		F-Test		
	Number of Websites	Cronbach's Alpha	Intraclass Correlation	Lower	Upper	Value	df	*p*
E.U. Nations								
Denmark	6	0.865	0.502	0.185	0.877	7.385	5	0.000
French	13	0.916	0.300	0.108	0.741	11.843	5	0.000
German	17	0.965	0.635	0.376	0.916	28.742	5	0.000
Netherlands	36	0.987	0.668	0.429	0.925	75.617	5	0.000
Spain	4	0.503	0.211	-0.124	0.757	2.013	5	0.135
Sweden	2	0.762	0.615	-0.180	0.934	4.200	5	0.071
UK	49	0.959	0.264	0.113	0.692	24.165	5	0.000
Non-EU nations with Comprehensive Privacy Legislation								
Australia	8	0.923	0.641	0.336	0.922	12.911	5	0.000
Canada	26	0.966	0.505	0.266	0.864	29.530	5	0.000
Japan	15	0.925	0.381	0.161	0.801	13.403	5	0.000
Nations without Comprehensive Privacy Legislation								
Korea	2	-4.800	-0.923	-1.335	0.152	0.172	5	0.962
Russia	20	1.000	1.000	1.000	1.000	625.000	5	0.000
U.S. Websites								
Safe Harbor	8	0.423	0.080	-0.050	5.350	1.734	5	0.153
US	415	0.996	0.342	0.168	0.759	257.050	5	0.000

circumstances such as for Korean websites in Table 18, Cronbach's alpha can become negative when the average covariance among items is negative [58].

The results in Tables 4-16 to 4-18 show little consistency in the pass/fail sequences at the national level (except Russia). However, this does not exclude the possibility of "clusters" of consistent behavior within the websites of a particular nation. To examine this possibility, we propose to treat these postulated clusters as latent categories, and employ a variant of exploratory factor analysis to identify these latent variables. Exploratory factor analysis [59] is often advocated as a plausible technique for identifying the latent structure of a population. However, standard factor analysis techniques are not applicable to our data, as they require both observed and latent variables to be continuous in nature (not categorical). The use of Tetrachoric correlation and Item Response Theory [60] have been proposed for latent trait analysis. These methods map dichotomous responses onto continuous latent variables; however, our interest is in discovering *classes* of websites with homogenous pass/fail sequences for our rules; thus, we seek to map dichotomous responses onto categorical latent variables.

Exploratory Latent Class Analysis (ELCA) [49, 61, 62] is analogous to exploratory factor analysis when the observed and latent variables are categorical. Through ELCA, observed response patterns composed of categorical responses are mapped onto latent classes utilizing a Bayesian statistical model and maximum likelihood estimation. This mapping of response patterns onto latent classes is analogous to exploratory cluster analysis. However, when we consider the number of dichotomous variables (7), we have 2^7=128 possible input combinations, with only 2172 total observations (for full policies). Since the Bayesian statistical model requires the usage of cross tabulation tables [62], there is a reasonable chance that at least some of these tables will be sparse. This invalidates the likelihood ratio Chi-Square goodness-of-fit test [62], which assumes that all response variables exceed a minimum frequency; such as the thresholds defined in [37]. Fitness testing for data such as ours is usually performed using either resampling approaches [63, 64], or heuristics [62, 65].

Resampling methods are usually preferred over heuristics for determining model fit. We will employ bootstrap resampling techniques since they are well suited for model testing [64]. In bootstrap resampling, a dataset is sampled with replacement, creating a pseudo-population. This process is repeated many times, generating a large number of pseudo-populations. The test statistic of interest (likelihood ratio Chi-Square test) is then measured for each pseudo-population. This approach provides unbiased estimates of the standard errors specific to the original dataset and allows for the creation of confidence intervals which can be employed to determine model fit [64]. If the test statistic for the original dataset lies outside the 95% confidence interval of the distribution of statistic values for the pseudo-populations, we reject the proposed model. If a threshold significance other than 0.05 is selected, the confidence interval used is adjusted accordingly. Since exploratory latent class analysis using resampling techniques is dependent upon the sizes of the observed populations, its use on populations other than the United States is questionable. For this reason, resampling methods were only employed for this population.

If resampling methods are inappropriate, heuristics may be used to judge model fit. Caution is however urged in the interpretation of these results unless further evidence is found indicating these results to be statistically significant, or they are found to be significant through meta-analysis techniques in further studies [38]. Two popular heuristics utilized in ELCA are Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) [65]. While some ELCA is completed through the use of only one heuristic, this practice is not recommended [65]. However, little guidance exists for how one should handle discrepancies between heuristics when two or more are utilized. Often the Cattell Scree test [66] is used; researchers visually determine where significant changes have occurred when moving from one latent model to another. The Scree test relies on the identification of a "leveling off" point that is assumed to indicate the point where the addition of further factors (or classes) leads to minimal model improvement.

We have used the LCAP tool [67] to generate our ELCA models. Table 4-19 indicates that the one latent class model could not be rejected for any of the U.S. populations under analysis. In using resampling techniques to judge ELCA model fitness, we do not move to a more complex model unless the current model is rejected; thus, resampling techniques leave us with the perfectly uninformative result of one latent class in the U.S. data. Since the application of resampling methods provided little insight, a heuristic based analysis using the Scree test and the AIC/BIC heuristics will be employed. The results of this analysis indicate that a two latent class model appears appropriate for the French (Figure 4-1), Dutch (Figure 4-11), Russian (Figure 4-19), and U.S (Figure 4-21) populations when the response patterns were developed by applying Rules 1-7 on the contents of full P3P policies. In all four of these populations, it can be observed that a 'significant change' occurred for the 'Full AIC' and 'Full BIC' lines in the various figures. Through a visual inspection of the conditional probabilities of the 2-latent class models for these populations it was found that French websites were classified based upon whether they passed or failed Rule 3. Dutch sites were classified based upon passing or failing Rule 6. Russian Websites were classified on whether the failed or passed both Rules 4 and 5, and finally U.S. websites were classified based upon whether they passed or failed Rules 2 and 6.

The only other nation which exhibited a 'significant change' in full policies was South Korea where a three latent class model appears to be appropriate. A three latent class

Table 4-19: Results of Exploratory Latent Class Analysis (Resampling)

Population	Number of Latent Classes	Total number of populations	Log likelihood Chi-Square	*p*	Decision
Both	1	1000	358.478	0.999	accept
Full	1	1000	1611.502	1.000	accept
Compact	1	1000	954.888	1.000	accept

model also appears appropriate for Korean websites posting P3P compact policies. Websites belonging to class 1 passed Rule 2, websites belonging to class 2 failed Rules 5 and 6 and websites belonging to class 3 passed Rule 3. Nothing can be said about the remaining populations due to the lack of obvious leveling off points in Figures 4-8 to 4-21. Since little similarity exists in the identified latent class models, no generalizations or conclusions can be made from these results other than to say that little uniformity appears to exist in the various populations under analysis.

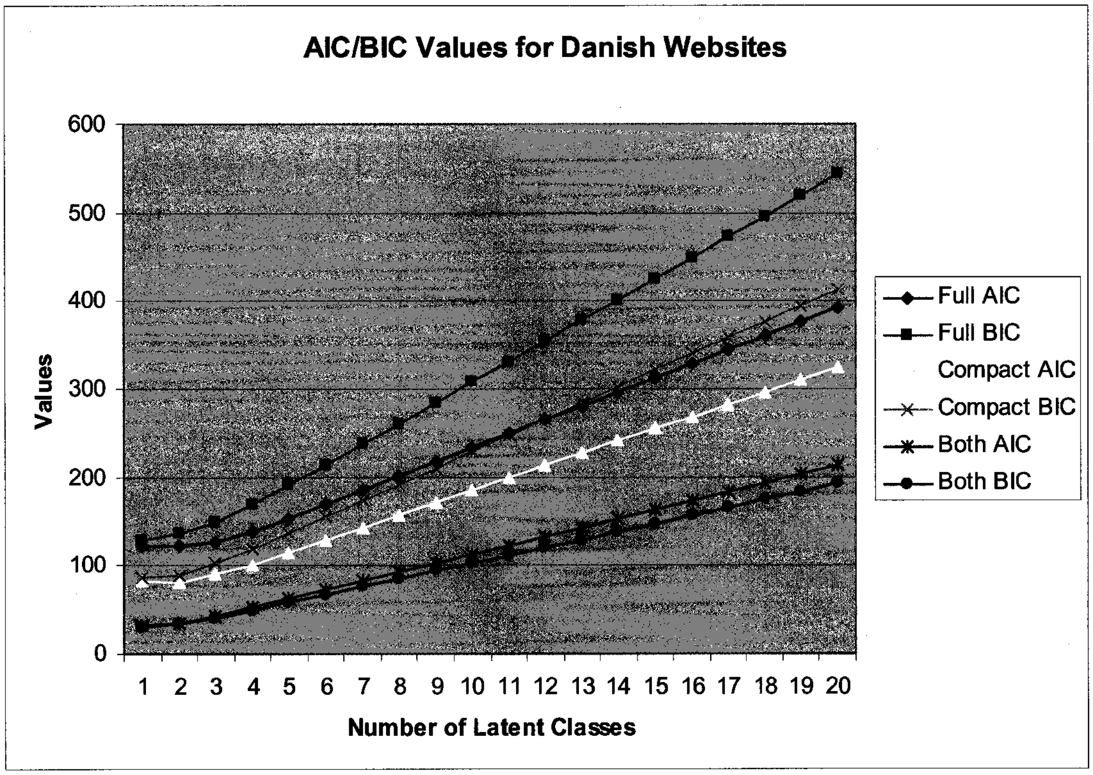

Figure 4-8: Intraclass Correlation Analysis for Danish Websites

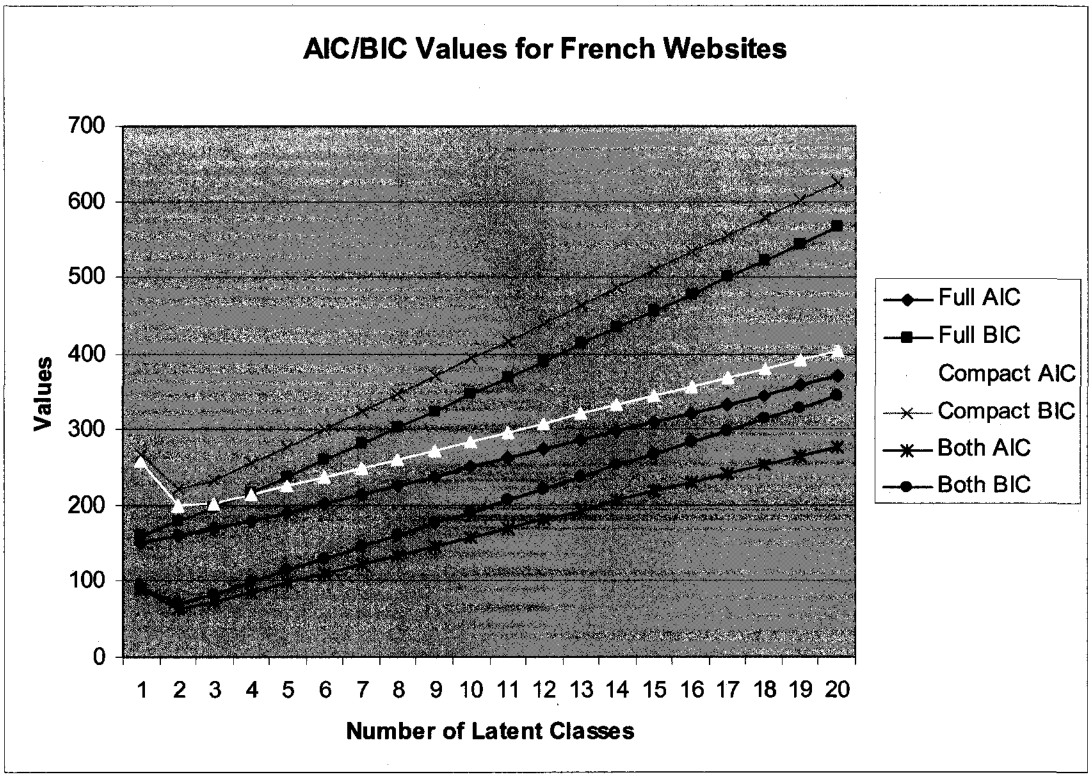

Figure 4-9: Intraclass Correlation Analysis for French Websites

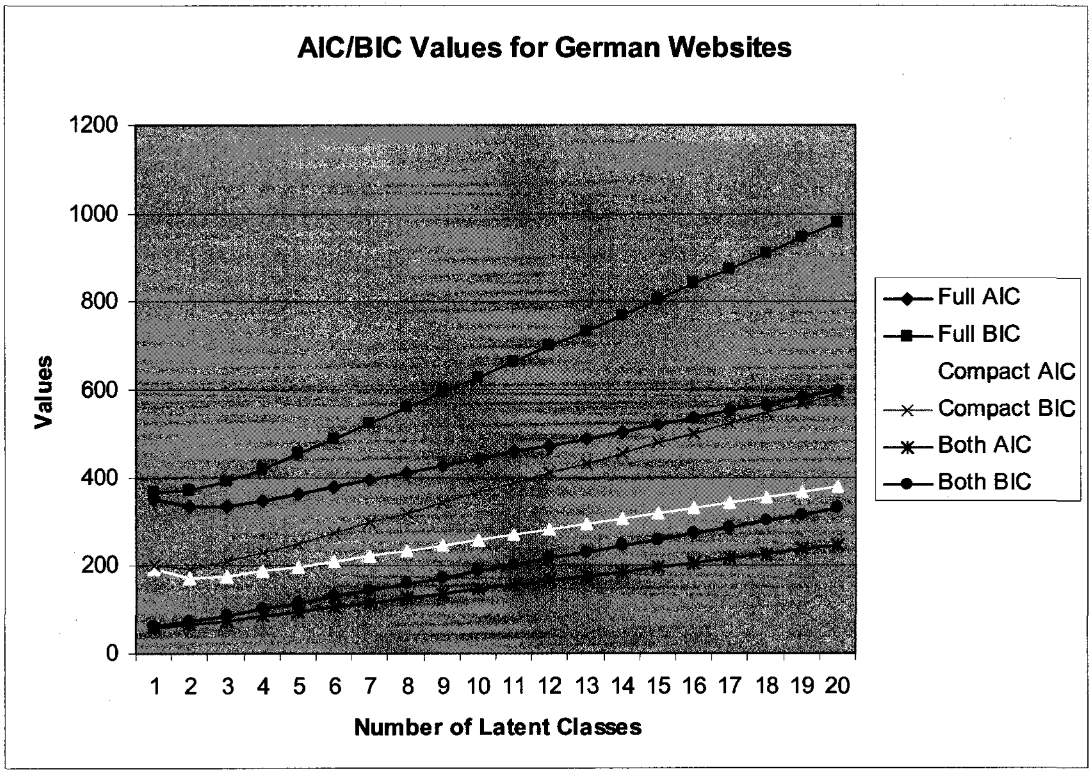

Figure 4-10: Intraclass Correlation Analysis for German Websites

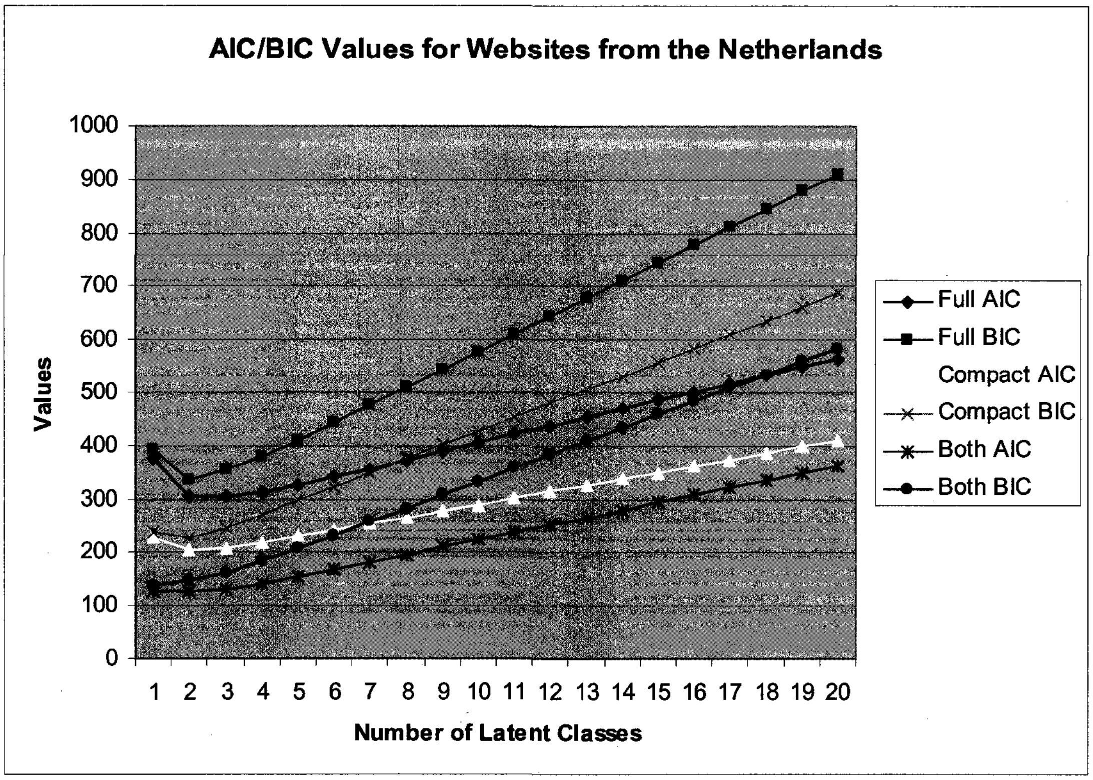

Figure 4-11: Intraclass Correlation Analysis for Dutch Websites

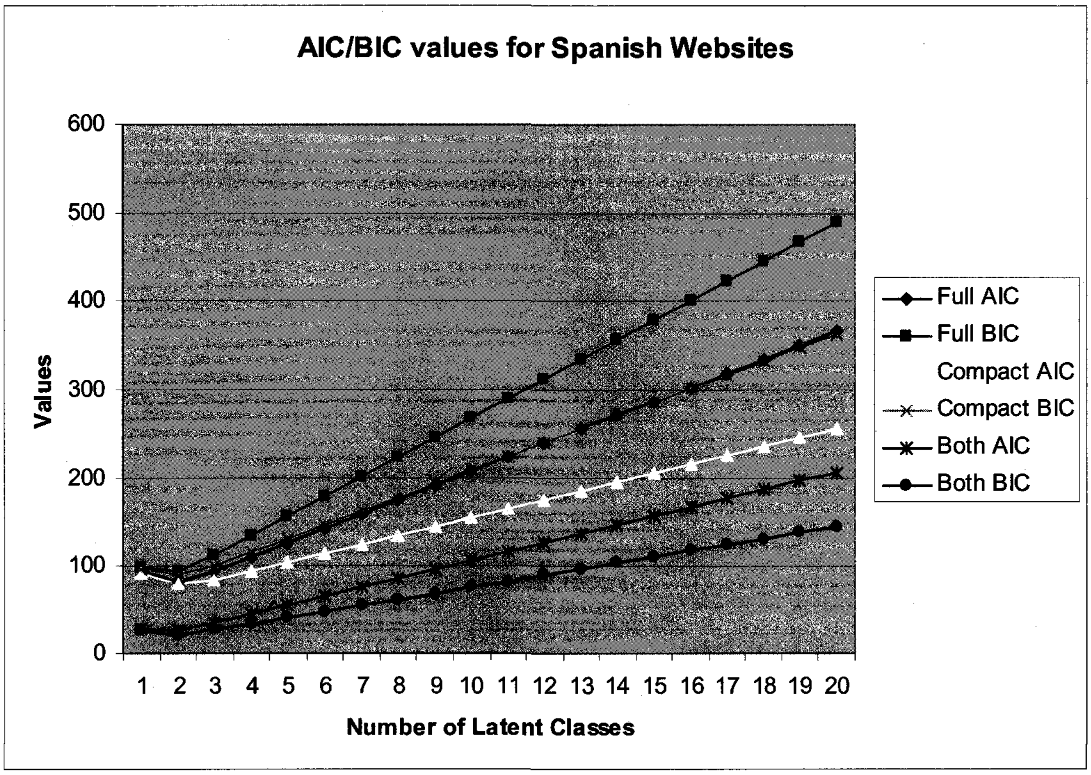

Figure 4-12: Intraclass Correlation Analysis for Spanish Websites

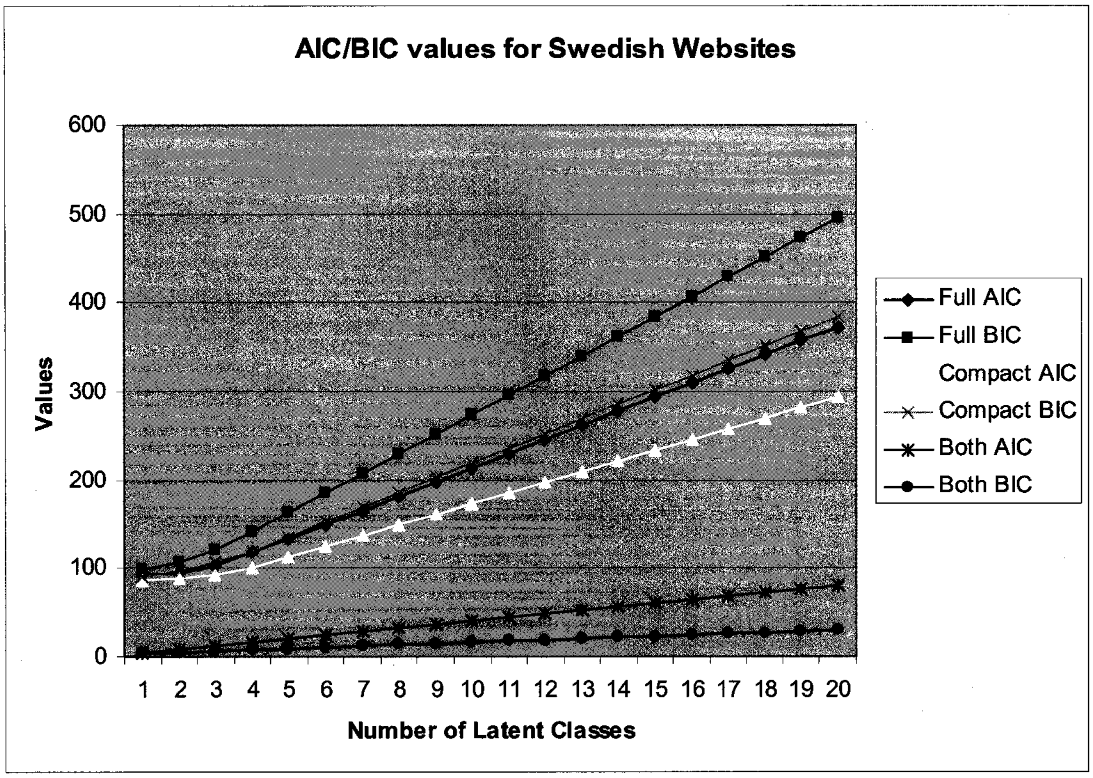

Figure 4-13: Intraclass Correlation Analysis for Swedish Websites

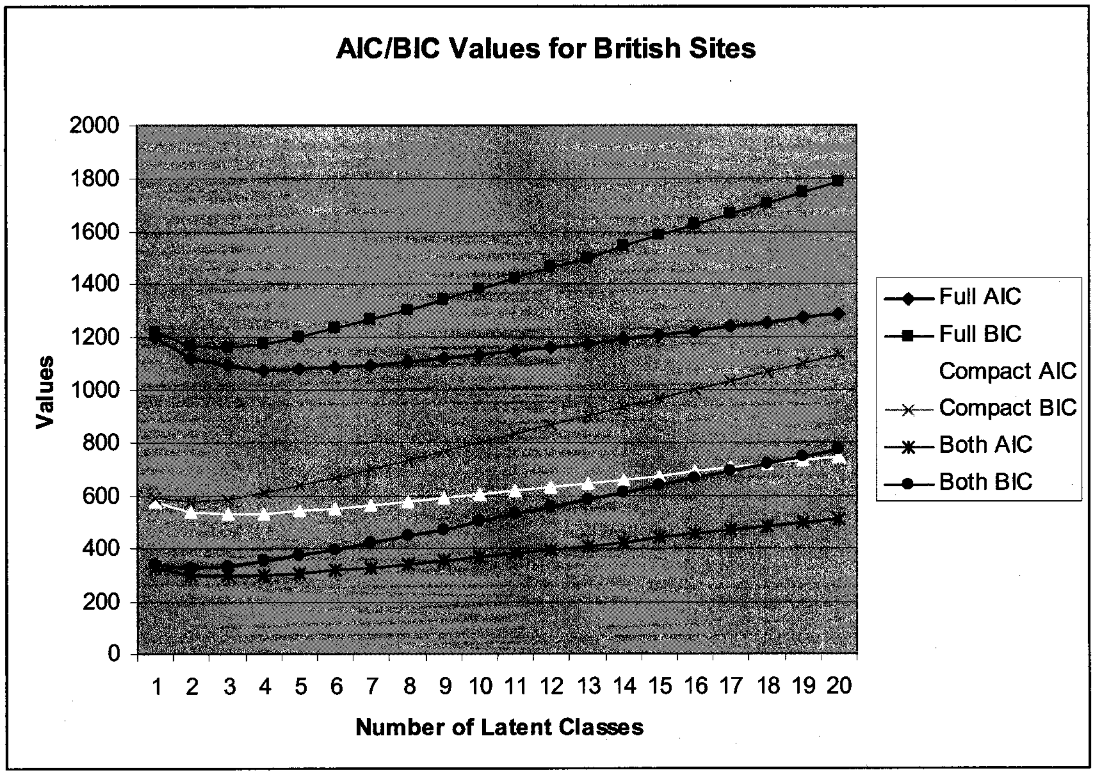

Figure 4-14: Intraclass Correlation Analysis for British Websites

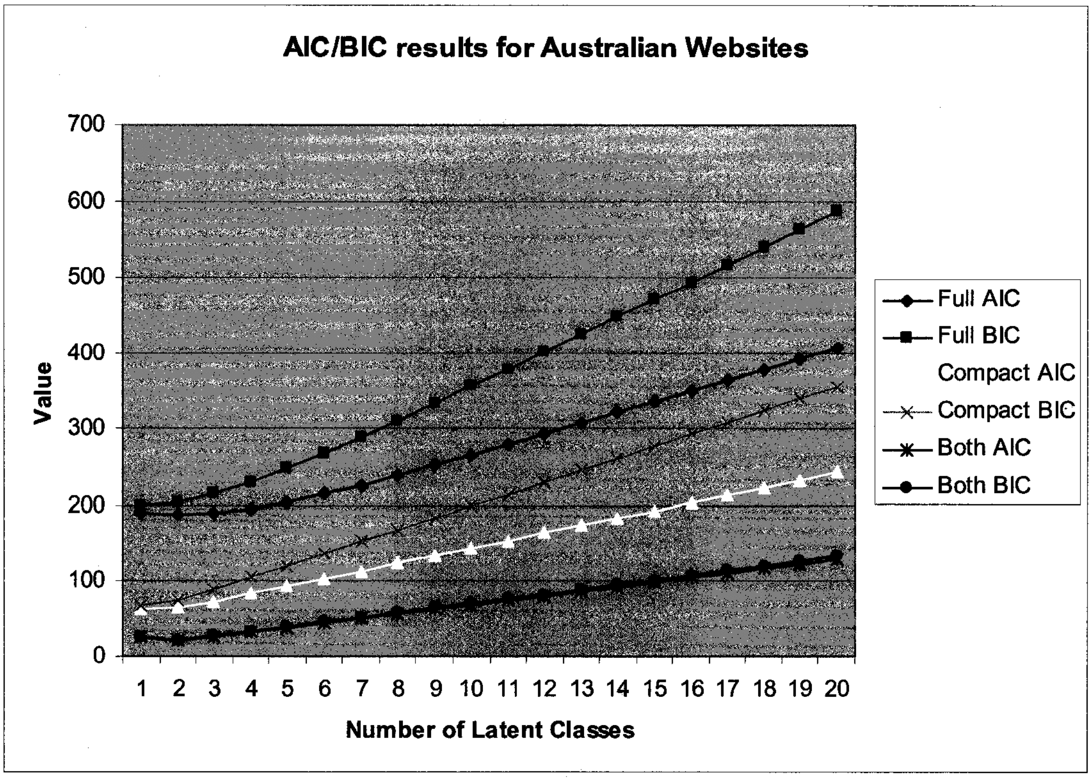

Figure 4-15: Intraclass Correlation Analysis for Australian

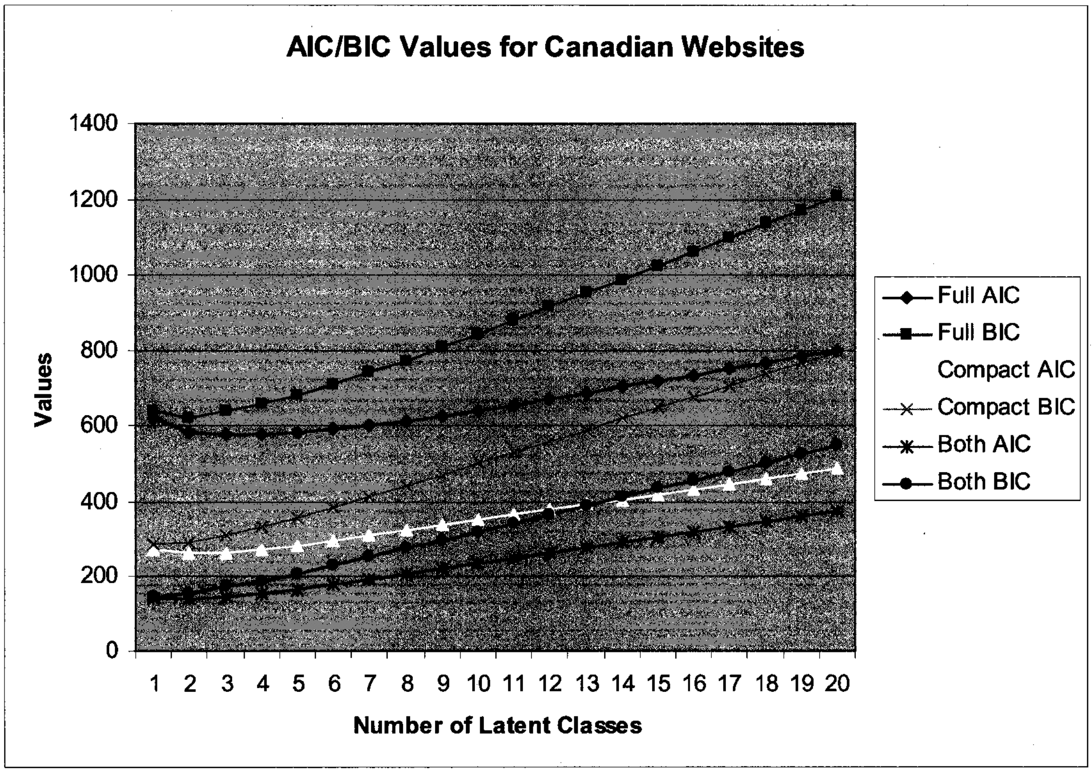

Figure 4-16: Intraclass Correlation Analysis for Canadian

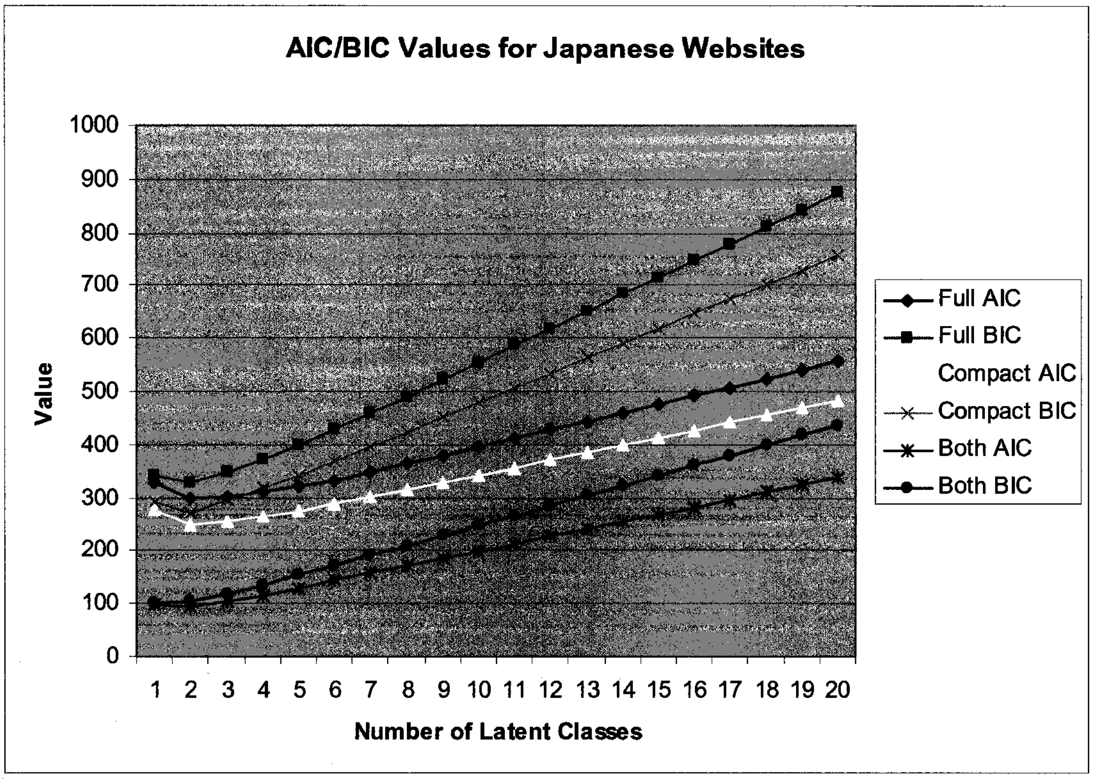

Figure 4-17: Intraclass Correlation Analysis for Japanese Websites

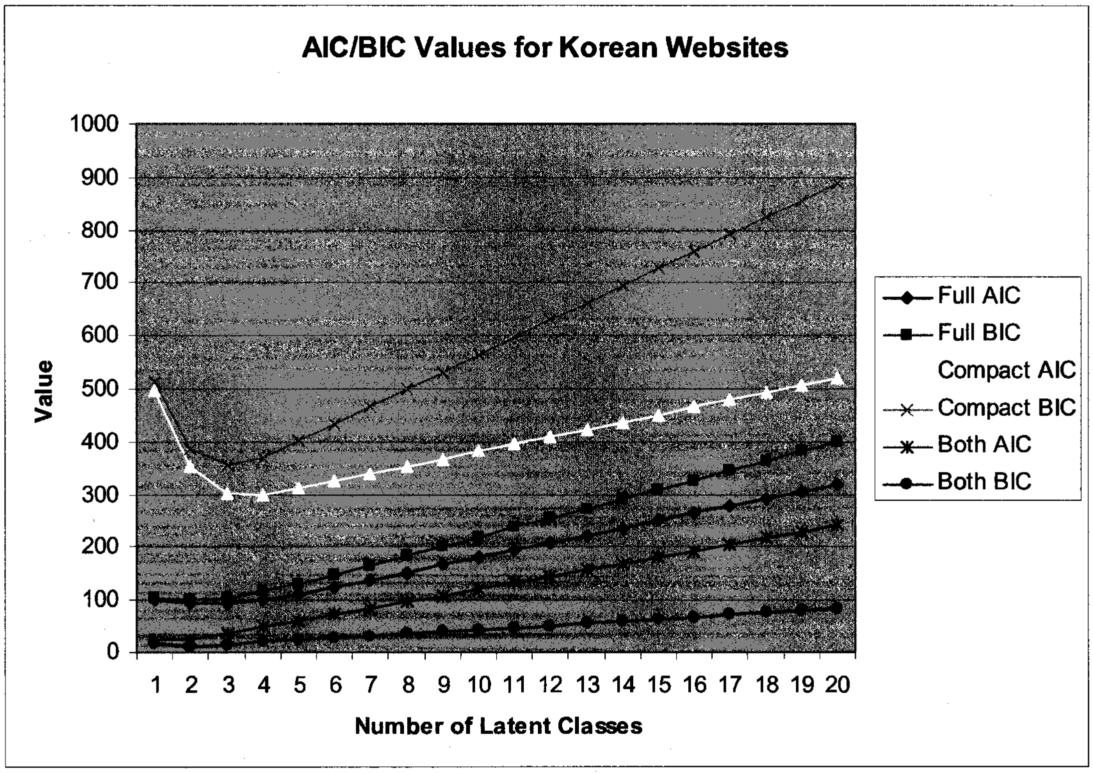

Figure 4-18: Intraclass Correlation Analysis for South Korean Websites

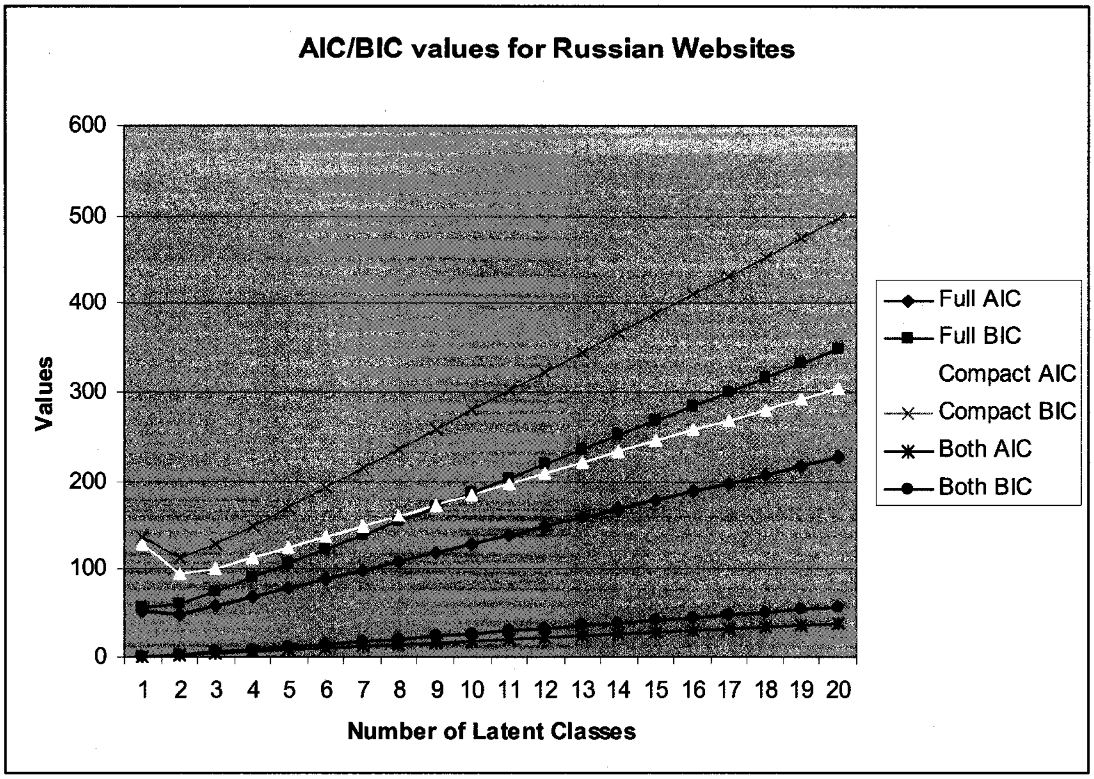

Figure 4-19: Intraclass Correlation Analysis for Russian Websites

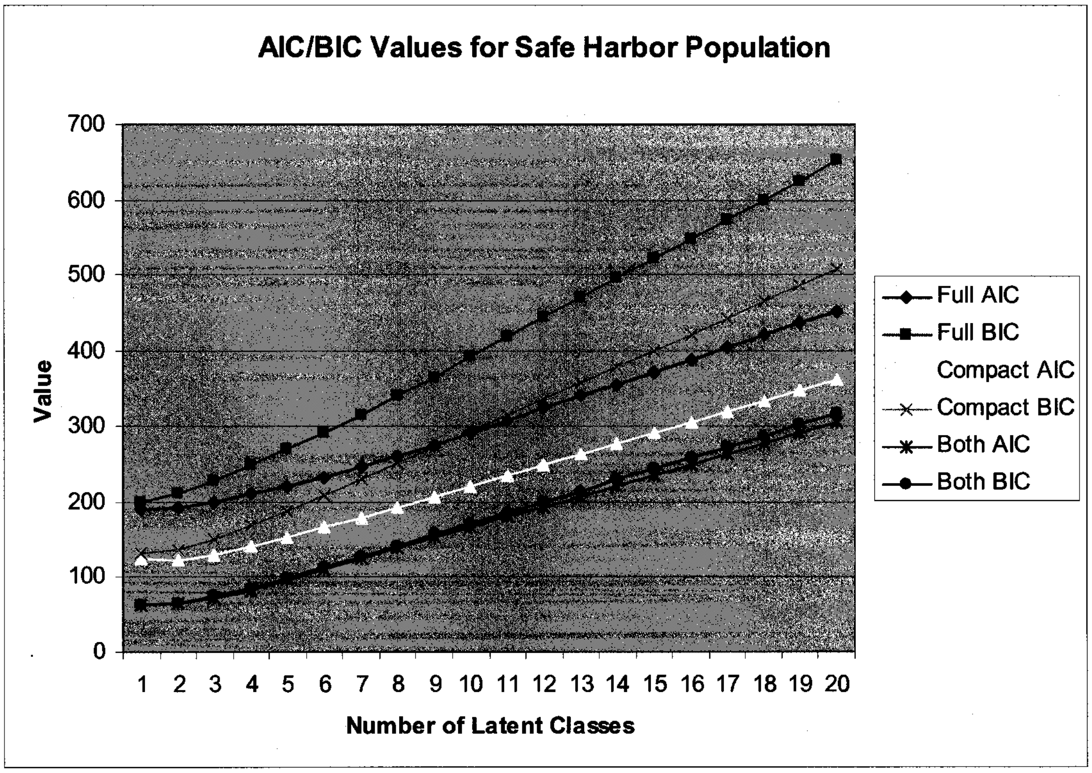

Figure 4-20: Intraclass Correlation Analysis for Safe Harbor Program

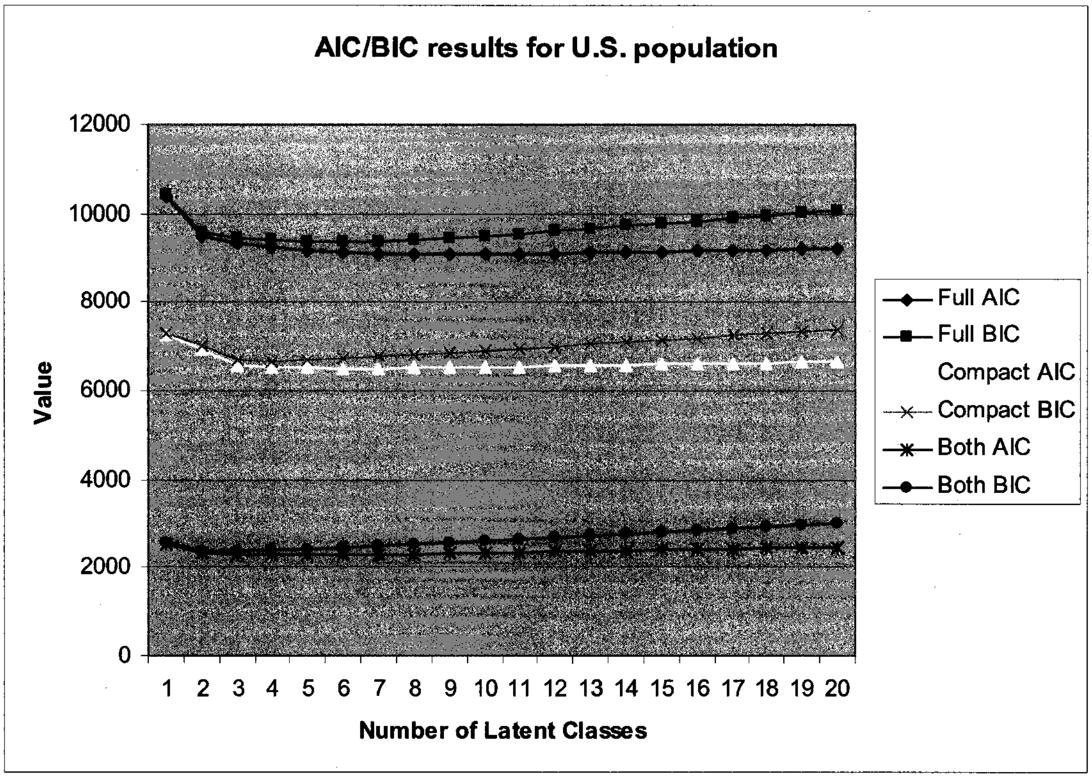

Figure 4-21: Intraclass Correlation Analysis for U.S. Websites

4.6 Conclusions and Future Work

In this paper, we attempt to expand upon the current work investigating the enforcement of privacy principles through legislation. This survey is unique since it utilizes the P3P protocol as a vehicle for automated data collection and analysis. This choice allows for a far more comprehensive analysis of stated website actions than otherwise could have been completed. The results of our analysis indicate a widespread lack of adherence to legal mandates in the stated actions of surveyed websites. This finding brings into question the effectiveness of comprehensive privacy protection laws such as the E.U. Data Protection Directive. Furthermore, adherence to government programs such as the U.S. Safe Harbor program was also found to be 'very poor'. In particular, rules governing the ability to access information and retention times for collected information were found to be 'poor'. In addition to the poor rates of adherence, statically significant differences were found between many E.U. nations, bringing into question any postulated "harmonizing" influence of the Data Protection Directive.

From these results, we conclude that the Data Protection Directive and Safe Harbor programs have had little effect upon the practices of websites analyzed in this paper. This raises the question of whether current attempts by the APEC ministers to create a comprehensive privacy framework similar to the Data Directive [6] will succeed. The results presented in this paper suggest that such attempts to create stringent harmonizing international privacy agreements have not been effective to date, and improvements will likely require a significant amount of time since the Data Protection Directive has now been in force for 12 years. We can only conclude that such attempts should be considered premature since our results indicate that additional research into the effectiveness of regulatory frameworks is required if policy makers are to avoid the current pitfalls which plague existing legislation.

Our results also bring into question attempts to develop environments such as the Safe Harbor Program. The lack of any statistically significant differences between Safe Harbor websites and the general U.S. population suggest that the self certification process has been largely ignored by surveyed websites. It seems that this problem may also extend to HRPP documents as well. Recent results presented in [68], show that HRPPs posted by nine organizations regulated by HIPPA (sectoral legislation for the U.S. health-care industry) have become more descriptive since HIPPA was introduced, but still provide consumers with little control over some of their most sensitive information. The documents are now also more difficult to comprehend. The self-regulatory approach to privacy protection in the U.S. does not seem to lead to strong privacy protections, but to an every-man-for-himself jungle.

Our results also suggest that inconsistencies in stated actions plague the websites of most nations under analysis. These inconsistencies bring into question whether Internet users can rely upon the stated actions of websites as a reliable indicator of their actual actions. There is, at this time, no mechanism by which users can easily verify the compliance of a data collector with the collector's posted privacy policy. Given the chaotic and contradictory environment we have observed in this study, that is a sobering thought. Research into the actual privacy-sensitive actions of websites, not just their public statements, is plainly needed.

This analysis also indicates that further work is required in determining why websites provide P3P policies. Our results depict various discrepancies between nations and between types of P3P policies. Given a chaotic environment, why do websites operators invest the resources to create a P3P policy in the first place? Without an enhanced understanding of the motivational factors behind P3P policy adoption, neither P3P adoption nor the protocol itself are likely to improve.

Our work does highlight a previously unexplored application of the P3P protocol as a remote auditing tool of website actions. If future extensions to the P3P protocol allow for a more accurate mapping between legal documents and P3P policies, P3P may prove to be invaluable to regulatory agencies since it would provide them an automated method of investigating websites. Such modifications may require P3P XML Schemas individually tailored to the legislation of individual jurisdictions. If such policies were to be developed, it is conceivable that tools such as the JRC Policy Workbench could be provided for websites of other. This would provide a convenient tool for website operators to craft appropriate policies for the jurisdictions they operate within.

4.7 Bibliography

[1] P. V. Lawson, Jeffery, "On the Data Trail: How Detailed Information About You Gets Into The Hands of Organizations With Whom You Have No Relationship," The Canadian Internet Policy and Public Interest Clinic, Ottawa, Ontario, Canada April 2006.

[2] J. L. Seligy, P., "Compliance with Canadian Data Protection Laws: Are Retailers Measuring Up?," The Canadian Internet Policy and Public Interest Clinic 2006.

[3] N. E. Bowie and K. Jamal, "Privacy Rights on the Internet: Self-Regulation or Government Regulation," *Business Ethics Quarterly,* vol. 16, July 2006 2006.

[4] M. Markel, "Safe Harbor and Privacy Protection: A Looming Issue for IT Professionals," *IEEE Transactions on Professional Communication,* vol. 49, pp. 1-11, 2006.

[5] R. C. Wenning, L., "The Platform for Privacy Preferences (P3P) Project." vol. 2007: World Wide Web Consortium, 2006.

[6] A. E. Chair, "APEC Privacy Framework," Asia-Pacific Economoic Cooperation, Santiago, Chile 2004/AMM/014rev1, 2004.

[7] OECD, "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data." vol. 2006: Organisation for Economic Co-operation and Development Directorate for Science, Technology and Industry, 1980.

[8] T. G. o. t. U. S. o. America, "Gramm-Leach-Bliley Act," Federal Trade Commission, 1999.

[9] T. G. o. t. U. S. o. America, "The Fair Credit Reporting Act." vol. 15 U.S.C. §

1681: Federal Trade Commission, 2004.

[10] T. G. o. t. U. S. o. America, "Children's Online Privacy Protection Act of 1998," T. G. o. t. U. S. o. America, Ed.: Federal Trade Commission, 1998.

[11] E. P. a. t. C. o. t. E. Union, "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data." vol. 95/46/EC E. P. a. t. C. o. t. E. Union, Ed.: Official Journal L 281

1995.

[12] T.-s. Parliament, "Personal Information Protection and Electronic Documents Act." vol. C-6 2000, T. G. o. H. M. t. Q. i. R. o. Canada, Ed.: Public Works and Government Services Canada - Publishing, 2000.

[13] Export.gov, "Safe Harbor Program." vol. 2007: U.S. Department of Commerce, International Trade Administration, 2007.

[14] L. F. Cranor, P. Guduru, and M. Arjula, "User Interfaces for Privacy Agents," *ACM Transactions on Computer-Human Interaction,* vol. 13, pp. 135 - 178, 2006.

[15] J. Staff, "JRC P3P Resource Center." vol. 2006: Joint Research Center, Ispra, 2006.

[16] L. L. Cranor, M.; Marchiori, M.; Presler-Marshall, M.; Reagle, J. , "The Platform for Privacy Preferences 1.0 Specification." vol. 2006: World Wide Web Consortium (W3C), 2002.

[17] L. F. Cranor, S. Byers, and D. Kormann, "An Analysis of P3P Deployment on Commercial, Government, and Children's Web Sites as of May 2003," AT&T Technical Report 2003.

[18] S. Egelman and L. F. Cranor, "An Analysis of P3P-Enabled Web Sites among Top-20 Search Results," in *Eighth International Conference on Electronic Commerce*, Fredericton, New Brunswick, Canada, 2006.

[19] I. K. Reay, P. Beatty, S. Dick, and J. Miller, "A Survey and Analysis of the P3P Protocol's Agents, Adoption, Maintenance and Future," *IEEE Transactions on Dependable and Secure Computing,* In press.

[20] J. B. Earp, A. I. Antón, L. Aiman-Smith, and W. H. Stufflebeam, "Examining Internet Privacy Policies Within the Context of User Privacy Values," *IEEE Transactions on Engineering Management,* vol. 52, pp. 227-237, May 2005.

[21] A. I. I. Staff, "Alexa Web Search - Top 500." vol. 2007: Amazon.com, 2007.

[22] I. L. c. Staff, "Geolocation IP Address to Country City Region Latitude Longitude ZIP Code ISP Domain Name Database for Developers." vol. 2007: IP2Location.com, 2006.

[23] H. Black, "On-Line Data Brokers." vol. 2006: Office of the Privacy Commissioner of Canada, 2005.

[24] D. Jutla and Y. Zhang, "Maturing e-Privacy with P3P and Context Agents," in *The 2005 IEEE International Conference on e-Technology, e-Commerce, and e-Services*, Hong Kong, 2005, pp. 556-561.

[25] D. Crocker and P. Overell, "Augmented BNF for Syntax Specifications: ABNF," The Internet Society 1997.

[26] L. Cranor, M. Langheinrich, and M. Marchiori, "A P3P Preference Exchange Language 1.0 (APPEL 1.0)." vol. 2006: W3C, 2002.

[27] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "An XPath-based preference

language for P3P," in *12th International Conference on World Wide Web*, Budapest, Hungary, 2003, pp. 629-639.

[28] G. Hogben, "A Technical Analysis of Problems with P3P v1.0 and Possible Solutions," Joint Research Centre November 12-13 2002.

[29] A. Watt, *Beginning Regular Expressions*. Indianapolis Indiana: Wiley Publishing, Inc., 2005.

[30] W. F. Adkinson, J. A. Eisenach, and T. M. Lenard, "Privacy Online: A Report on the Information Practices and Policies of Commercial Web Sites," The Progress and Freedom Foundation 2002.

[31] a. t. S. a. t. H. o. R. o. t. C. o. A. The Queen, "Privacy Act 1988." vol. 1988-Act 119, C. o. Australia, Ed.: Office of Legislative Drafting and Publishing, Attorney-General's Department, Canberra, 1988.

[32] P. I. Staff, "PHR2004 - Japan." vol. 2006: Privacy International, 2004.

[33] P. I. Staff, "PHR 2004 - The Russian Federation." vol. 2006, PrivacyInternational.org, Ed.: PrivacyInternational.org, 2004.

[34] M. T. S. Rajan, "The Past and Future of Privacy in Russia," *Review of Central and East European Law,* vol. 27, pp. 625-638, 2002.

[35] P. I. Staff, "PHR2004 - The Republic of Korea." vol. 2006: Privacy International.org, 2004.

[36] J. A. Rice, *Mathematical Statistics and Data Analysis*, 3rd ed. Belmont, CA: Thompson/Brooks/Cole, 2007.

[37] D. J. Sheskin, *Handbook of Parametric and NonParametric Statistical Procedures*, 3rd ed. Boca Raton: Chapman and Hall/CRC, 2004.

[38] M. W. Lipsey and D. B. Wilson, *Practical Meta-Analysis.* Thousand Oaks, California: Sage Publications, 2001.

[39] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences (2nd Edition).* Hillsdale, NJ: Lawrence Earlbaum Associates, 1988.

[40] V. Hasselblad and L. V. Hedges, "Meta-Analysis of Screening and Diagnostic Tests," *Psychological Bulletin,* vol. 117, pp. 167-178, January 1995.

[41] J. Copas and D. Jackson, "A Bound for Publication Bias Based on the Fraction of Unpublished Studies," *Biometrics,* vol. 60, pp. 146-153, March 2004.

[42] G. o. Sweden, "Personal Data Act." vol. 1998:204, G. o. Sweden, Ed.: The Swedish Data Inspection Board, 1998.

[43] S. Bellman, E. J. Johnson, and G. L. Lohse, "On site: to opt-in or opt-out?: It Depends on the Question," *Comunications of the ACM,* vol. 44, pp. 25-27, 2001.

[44] P. I. Staff, "PHR2004 - Federal Republic of Germany." vol. 2006: Privacy International, 2004.

[45] C. R. Française, "Comission Nationale De L'Informatique Et Des Libertes." vol. 2006: Comission Nationale De L'Informatique Et Des Libertes, 2006.

[46] L. J. Cronbach, "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika,* vol. 16, pp. 297-334, 1951.

[47] G. D. Garson, "PA 765: Scales and Standard Measures," in *PA 765: Multivariate Analysis in Public Administration.* vol. 2007, P. G. D. Garson, Ed.: College of Humanities and Social Sciences, North Carolina State University, 2007.

[48] J. C. Nunnally, *Psychometric Theory.* New York: McGraw-Hill Inc., 1978.

[49] E. B. Andersen, *The Statistical Analysis of Categorical Data.* New York:

Springer-Verlag, 1990.

[50] D. C. Howell, *Statistical Methods for Psychology*. Pacific Grove, CA, USA: Duxbury/Thomson Learning, 2002.

[51] L. G. Portney and M. P. Watkins, *Foundations of Clinical Research, Applications to Practice*. Norwalk, CN, USA: Appletop & Lange, 1993.

[52] P. E. Shrout and J. L. Fleiss, "Intraclass Correlations: Uses in Assessing Rater Reliability," *Psychological Bulletin,* vol. 86, pp. 420-428, 1979.

[53] J. C. Nunnally and I. H. Bernstien, *Psychometric Theory*. New York: McGraw-Hill, Inc., 1994.

[54] S. D. Walter, M. Eliasziw, and A. Donner, "Sample size and optimal designs for reliability studies," *Statistics in Medicine,* vol. 17, pp. 101-110, 1998.

[55] H. S. Bloom, "Minimum Detectable Effects," *Evaluation Review,* vol. 19, pp. 547-556, 1995.

[56] D. G. Bonett, "Sample Size Requirements for Estimating Intraclass Correlations with Desired Precision," *Statistics in Medicine,* vol. 21, pp. 1331-1335, 2002.

[57] I. K. Reay, P. Beatty, S. Dick, and J. Miller, "Privacy Policies Versus National Culture and Legislation on the Internet," *ACM Transactions on Computer-Human Interaction,* Submitted.

[58] D. P. Nichols, "My Coefficient alpha is Negative!." vol. 2006: SPSS Inc., 1999.

[59] B. Thompson, *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*. Washington, D.C, USA: American Psychological Association, 2004.

[60] R. J. Mislevy, "Recent Developments in the Factor Analysis of Categorical

Variables," *Journal of Educational Statistics,* vol. 11, pp. 3-31, 1986.

[61] A. Agresti, *Categorical Data Analysis*. New York, NY: Wiley-Interscience, 2002.

[62] A. L. McCutcheon, *Latent Class Analysis*. Newbury Park, CA, USA: Sage Publications, Inc., 1987.

[63] J. G. Dias and J. K. Vermunt, "Bootstrap Methods for Measuring Classification Uncertainty in Latent Class Analysis," in *Proceedings in Computational Statistics*, A. Rizzi and M. Vichi, Eds. Heidelberg: Springer, 2006, pp. 31-41.

[64] R. Langeheine, J. Pannekoek, and F. V. D. Pol, "Bootstrapping Goodness-of-Fit Measures in Categorical Data Analysis," *Sociological Methods and Research,* vol. 24, pp. 492-516, 1996.

[65] J. Kuha, "AIC and BIC Comparisons of Assumptions and Performance," *Sociological Methods and Research,* vol. 33, pp. 188-229, 2004.

[66] R. B. Cattell, "The Scree Test for the Number of Factors," *Multivariate Behavioral Research,* vol. 1, pp. 245-276, 1966.

[67] R. Neuman, "LCAP 2.33." vol. 2007, 2003.

[68] A. I. Anton, J. B. Earp, M. W. Vail, N. Jain, C. M. Gheen, and J. M. Frink, "HIPPA's Effect on Web Site Privacy Policies," *IEEE Security and Privacy,* vol. 5, pp. 45-52, 2007.

An Economic Model for Privacy Signals

5 Chapter 5 Introduction

On the commercial playing field, buyers and sellers (or their agents) come to a negotiation with different needs, expectations, and background knowledge. The classic example of this occurs in the used-car sales lot; buyers must differentiate between high quality (peach) and poor quality (lemon) cars. The seller has the advantage; they possess private information about the quality of their car. Buyers, in contrast, are at a disadvantage because they lack access to this information. This is an example of a transaction mired in *information asymmetry* [1].

Akerlof's celebrated work attempts to analyze markets where *information asymmetry* exists [1]. Akerlof's conclusion that markets suffering from asymmetric information will generally produce sub-optimal results for all transacting parties, stems from the perceived 'riskiness' of transactions within these markets. Risk adverse parties tend to either avoid these markets or demand significant incentives to take part.

Economists have identified three market problems caused by asymmetric information: *Moral Hazard*, *Adverse Selection*, and *Signaling* problems.

- *Moral Hazard* problems arise when some unobservable action presents an opportunity for a transacting party to benefit by cheating after the deal has been signed. Insurance fraud is an example of such a problem. Often it is difficult if not impossible for insurance companies to investigate claims in sufficient detail to discover whether the client is acting in good faith.
- *Adverse Selection* problems arise when one transacting party cannot verify attributes of another party before committing to a deal. The health insurance

industry suffers from this problem since it can be difficult to accurately assess the health of potential clients even with extensive examinations.

- *Signaling* problems are a special instance of the *Adverse Selection* problem, where a secondary piece of information (the signal) is used to indirectly indicate the presence or absence of an attribute. The usage of brand names in car vehicle markets is an example. Users cannot directly observe vehicle reliability and instead, use a vehicles brand name along with their own personal experiences as an estimate.

We propose that the Internet's privacy problems are an example of a *signaling* problem. Privacy conscious Internet users need to assess whether a website respects their patron's privacy or not, however they are often unable to access information needed to make an informed decision. For example, it is unlikely that an average Internet user would be able to access information describing which third party organizations a website will share collected information with [2] since Privacy policies may have their content obscured with complex legal terminology and deceptive statements [3]. P3P attempts to solve these issues by providing website patrons with signals derived from information contained in a P3P policy. Furthermore, the existence of a P3P policy, just like the existence of a human readable privacy policy, could be interpreted by users as a signal that the website is trustworthy [4].

In this paper, we propose to investigate the effectiveness of the P3P protocol as an online signaling mechanism. Our analysis will extend Spence's classic signaling model [5] to fit the online marketplace. We will then test it against information collected from various websites operating in distinct online markets. The remainder of this paper is organized as follows: In Section 5.1, we discuss the classic signaling model in the context of online privacy issues. In Section 5.2, we propose our economic model of privacy signals. In Section 5.3, we present our survey method. In Section 5.4, we present our survey results. In Section 5.5, we propose an extension of our model. Finally, in Section 5.6, we present our conclusions.

5.1 Signaling Theory and Online Privacy

Signaling theory assumes that markets are composed of two types of actors, those who request services (principles) and those who provide services (agents). In online markets, these roles are filled by websites (agents) and their patrons (principles). Signaling theory also assumes that subcategories of agents and principles (peach or lemon) exist within a market. The work of Ackerman *et al.* [6] has identified the following types of principles:

- Privacy fundamentalists who cherish their privacy and take significant steps to protect it.
- Pragmatic majority who are concerned about their privacy but less so than fundamentalists.
- Marginally concerned who share information with little concern for its use.

In our model, we propose to distinguish between principles willing to pay to protect their privacy and those who do not. Such a dichotomization appears reasonable given that the privacy fundamentalist group is closely associated with those willing to pay to protect their privacy. It also appears likely that a close association exists between the marginally concerned and those unwilling to pay to protect their privacy. Finally, the pragmatic majority would fall into one of these two groups depending upon the threat. We also propose that agents can be classified into dichotomous groups; those who respect privacy and those that do not. This dichotomization is supported by previous research [7] clearly indicating that websites differ significantly in how they use information.

Signaling theory proposes that the proportion of the marketplace occupied by peach/lemon principles and agents will eventually reach steady state equilibrium in the absence of external stimuli. This assumption is certainly not without problems. No marketplace is truly independent from the effects of other markets, the environment, or government actions. However, it is a common and necessary simplification, which still produces models that can still lead to significant insight into the effect signals have upon marketplaces [1, 5, 8]. Classic signaling theory proposes that two flavors of equilibria (separating or pooling) can develop within a marketplace. Before we compare and contrast these equilibria, an overview of the signaling process is required to understand why they develop.

The signaling process is characterized by continuously repeating three steps:

- **Step one:** agents invest in "objects" whose characteristics act as signals of an agent's type. In Spence's signaling model [5], agents invest in education whose characteristics (school name, field of study, grades, etc.) are treated as indicators of their competence. Instead of education, websites invest in objects, such as P3P policies, whose characteristics (existence of the policy, contents, comprehensiveness, etc.) are intended to indicate the website's type.
- **Step two:** principles actively test for the existence of these objects and interpret their characteristics. Based upon this interpretation, principles will develop an impression of the agents' type. For P3P, this involves P3P agents requesting policies and then comparing their contents to the user's stated privacy

preferences.

- **Step three:** principles decide whether they will transact with the agent. Assuming the principle chooses to transact with the agent, the outcome of the transaction will either reinforce or contradict the principles pre-existing beliefs as to what the signal implies. For instance, if a principle's privacy is violated by a website whose P3P policy passes their predefined preferences, the principle may modify their preferences to detect similar problems in the future. If principles begin to believe that P3P policy contents are not a reliable indicator of actual website actions, they may begin to believe that the P3P protocol as a whole is unreliable.

Assuming that principles believe the signal is a reliable indicator of agent type, principles will begin to categorize agents using these signals. When the membership of these categories stabilize – an equilibrium develops. Molho [9] defines two characteristics of signaling equilibria:

1) Agents have no incentive to change their signaling decision given the cost of signaling on one hand and the benefits received from being considered of a particular type of the other hand
2) Principles beliefs, which are based upon interpretations of signals, are confirmed.

As previously mentioned, signaling theory proposes that one of two possible equilibria (pooling or separating), will develop in markets relying upon signals. Pooling equilibria develop when principles logically pool agents of different types into a single common type. Principles are forced into this action when: no agent signals, all agents signal, or the signal is unreliable. In all three possibilities, the marketplace degrades into what Akerlof refers to as a “lemons market”.

Separating equilibria arise when signals allow principles to logically separate agents of different types, allowing principles to select agents with desirable traits. Such an equilibria can only be maintained if the principles’ interpretations of object characteristics are confirmed. These two equilibria are also examples of an additional form of equilibria, Nash equilibria, which is used extensively in game theory. When a Nash equilibria exists, no actor can increase their overall utility from a change in strategy without another actor also changing their strategy. These equilibria are often referred to as strategically stable because no incentive exists to motivate actors into changing their strategy. Molho [9] proposes that both pooling and separating equilibria are strategically stable. P3P was designed to provide a mechanism for creating separating equilibria and it is the goal of the remainder of this paper to explore whether this has occurred.

5.2 Economic Model for Privacy Signals

Signaling models rely upon utility functions to model the benefits realized by principles and their agents. Naturally, these models assume that principles and agents will attempt to maximize their utility gained from transactions they undertake. We shall use the utility function $U_A()$ to represent the overall utility realized by agents from a transaction and $U_P()$ to represent the overall utility realized by principles. Furthermore, we assume that the overall utility of a transaction realized by agents and principles is the difference between the benefits derived from the transaction ($B()$) and the disutility associated with expending time or resources to complete the transaction ($V()$).

Since principles enter into transactions with the goal of attaining some product or service (x), the benefits they receive from the transaction is dependent upon the product ($B(x)$).

Furthermore, when principles contract agents for goods or services, principles suffer a disutility in proportion to the payoff (*p*) that they provide to the agent ($V(p)$). This logic leads to the development of Equation 5-1:

Equation 5-1
$$U_P = B(x) - V(p)$$

Similar logic also holds regarding the overall utility experienced by agents which is the difference between the benefits derived from the principles payoff ($B(p)$) and the disutility arising from expending effort/resources (*e*) to provide goods or services ($V(p)$):

Equation 5-2
$$U_A = B(p) - V(e)$$

Since rational principles and agents will not enter into transactions where a negative utility is expected, it is assumed that both U_A and U_P are positive. In order to model the intangible benefits of having an individual's privacy protected, it is assumed that the benefits and disutility experienced by agents and principles are the result of both tangible and intangible benefits (*x*), payoffs (*p*), and effort (*e*).

From Equations 5-1 and 5-2, it can be easily observed that principles and agents maximize their transactional utility at the expense of each other. We propose that websites can increase their utility through the usage of information collected from principles. Significant evidence exists to support this assumption [2]. However, websites do not always disclose this information, creating an environment mired in information asymmetry. Since principles may not be informed as to how their information will be used, they loose control over the information, creating the opportunity for privacy violations. We choose to model this transfer of information control as part of the payoff principles give to agents in return for services rendered. Thus, the payoff agents receive is a combination of monetary benefits (d) and information (i) that can be used by agents to enhance their utility resulting in two possible payoffs for agents:

- $p_2 = d_2$ when websites signal they respect privacy; and
- $p_1 = d_1 + i$ when they fail to respect.

Since rational agents will only signal if they believe the action will result in increased utility, it is implied that $B(p_1) \leq B(p_2)$, without this condition, no rational agent would expend the effort to differentiate themselves. Similarly, no rational principle will provide the payoff p_2 without some form of assurance that their privacy will be respected. Hence, the signal must be perceived by principles as a reliable indicator of agent type. Otherwise, rational principles will assume that agents will adopt the signal no matter their type, collect the greater monetary payment (p_2), and still invade their patrons' privacy, since such actions would result in the greatest overall utility.

Signaling theory proposes that principles develop impressions of signal reliability from sources such as their own personal experiences, acquaintances, and those of reliable strangers. These impressions will naturally vary from principle to principle and we choose to model this perception through the inclusion of the function $q(t)$ where t is the signal observed by the principle, and $q()$ represents the user's perception of the signal's reliability. Since the payoff principles provide to agents is dependent upon their perception of the agent type, p will depend upon whether the website is able to persuade patrons that they respect their privacy, which in turn, depends upon the reliability of the signal. We choose to model this relationship through the addition of a threshold value z. If the website presents a signal which exceeds the users threshold ($q(t) \geq z$), it is assumed that the principle will believe that the agent will respect their privacy. Similarly, if $q(t) < z$, it is assumed that the principle will believe that the website will fail to respect their privacy. When taken together, these two assumptions establish:

Equation 5-3

$$p = \begin{cases} p_2 & if \quad q(t) \geq z \\ p_1 & if \quad q(t) < z \end{cases}$$

Thus, the utility gained by agents in the marketplace is:

Equation 5-4

$$U_A = \begin{cases} B(p_2) - V(e) & if \quad q(t) \geq z \\ B(p_1) - V(e) & if \quad q(t) < z \end{cases}$$

In the current model, websites do not have to invest additional resources to signal and thus receive a greater benefit since e is constant. If a separating equilibrium is to develop, only desirable (privacy respecting) websites should display signals that pass the principles quality threshold ($q(t) \geq z$). Implied in this statement is a requirement that the signal adoption cost should be low for respecting agents and high for disrespecting agents. If the adoption costs for disrespecting agents exceed the benefits they derive from masquerading as a respecting site, no rational disrespecting agent will adopt the signal establishing a separating equilibrium. Signal adoption costs could take the form of penalties enforced by a third party, public institutions, or the effort required to attain the object with the required characteristics. In this model, $c_R(t)$, and $c_D(t)$ represent the signaling cost borne by privacy respecting and disrespecting agents respectively. These costs create additional disutility for websites and leads to four possible combinations of agent utility (Equation 5-5):

Equation 5-5

$$U_A = \begin{cases} U(p_2) - V(e) - V(c_R(t)) & \text{if} \quad q(t) \geq z \quad \text{and} \quad \text{respects} \quad \text{privacy} \\ U(p_2) - V(e) - V(c_D(t)) & \text{if} \quad q(t) \geq z \quad \text{and} \quad \text{disrespects} \quad \text{privacy} \\ U(p_1) - V(e) - V(c_R(t)) & \text{if} \quad q(t) < z \quad \text{and} \quad \text{respects} \quad \text{privacy} \\ U(p_1) - V(e) - V(c_D(t)) & \text{if} \quad q(t) < z \quad \text{and} \quad \text{disrespects} \quad \text{privacy} \end{cases}$$

Based upon the above argument, signals only become reliable and the signaling equilibria are established, if $V(c_D(t)) > V(c_R(t))$ and the difference is sufficient to ensure that it is not in a disrespecting websites interest to invest $V(c_D(t))$ in an attempt to attain a signal that satisfies $q(t) \geq z$. If we assume that these preconditions have been met, Equation 5-5 can be greatly simplified. When separating equilibria develop, only those websites who provide a signal satisfying $q(t) \geq z$ will be rewarded by principles with the payoff p_2. All other agents, whether they respect privacy or not, will be rewarded with payoff p_1. Furthermore, since the signal is assumed to be reliable, only respecting websites will adopt the signal. These simplifications lead to the development of Equation 5-6:

Equation 5-6

$$U_A = \begin{cases} U(p_2) - V(e) - V(c_R(t)) & \text{if} \quad q(t) \geq y \quad \text{and} \quad \text{respects} \quad \text{privacy} \\ U(p_1) - V(e) & \text{otherwise} \end{cases}$$

A subtle assumption of this model is that principles will always test. In many marketplaces, this is likely to be a reasonable assumption since the cost of testing is very low when compared to the cost of choosing the wrong agent. For instance, it costs very little for a company to verify the credentials of potential employees when compared to the consequences of hiring a poor candidate. The web differs from these more standard applications of signaling because the cost to test can become significant when compared to the benefits received by the principle. Many web agents offer their services for free to individuals and the services offered are of relatively low benefit to customers. For example, the benefit gained by the user from reading their news online in contrast to reading it offline would amount to a relatively trivial sum for many individuals. To model testing costs, we propose to modify Equation 5-1 to include the cost of testing, resulting in Equation 5-7 where T represents the cost born by Principles who test.

Equation 5-7

$$U_P = B(x) - V(p) - V(T)$$

If the combination of the pay-off plus the cost to test becomes larger than the benefit received by the principle ($B(x) < V(p) - V(T)$), the transaction will no longer generate value for the principle leading to principles avoiding these transactions or simply 'taking their chances' and electing not to test. In such an environment, websites will not be provided with the payoff, p_2, since the principles will never view the signal. Instead, principles will logically pool websites together and provide a common payoff to all websites (p_1) resulting in privacy respecting websites suffering an economic disadvantage when compared to their disrespecting counterparts, leading to their eventual expulsion from the marketplace. It can be shown that a similar result will develop in all marketplaces where pooling equilibria develop [9]. Thus, when the preconditions for separating equilibria fail to hold, it can be assumed that pooling equilibria will develop.

From this model, it is readily apparent that two preconditions must be met for privacy signals to be effective:

- First, principles must actively test for the existence of the signal
- Second, principles must provide payoffs to desirable agents that lead to the agent experiencing an overall increase in utility.

Unfortunately, the results from our survey seem to imply that these preconditions are currently not met; and hence a market for the P3P protocol is still to fully form on the Internet.

5.3 Survey Method

We propose to test the validity of our signaling model for the P3P privacy signal by testing the first precondition for the establishment of separating equilibria; the active testing for signals by principles. If it is found that principles are actively testing for P3P policies, we will continue with a survey analyzing whether or not price differentials are being provided. This approach has been effective in investigating similar signaling mechanisms such as E-Bay's reputation system [10, 11] where it was identified that E-Bay's reputation mechanism significantly reduces the payoffs principles give to 'lemon' agents. The major difference between these surveys and ours is that we must first identify whether or not principles are actively testing for the signal. The E-Bay surveys assumed that principles were actively testing for the signals since they provide feedback in upwards of 50% of transactions [11].

In order to satisfy the first step of this process, we propose to analyze web server logs provided by several organizations to determine whether principles, or software agents working on their behalf, ever attempt to retrieve P3P policies. We have collected web server access logs from three organizations, which occupy diverse online market segments. The first organization, which will be referred to as 'org1' is a large public post-secondary academic institution. The server logs represent the web requests submitted to a range of individual websites hosted from a central server and represent a diverse set of topics. Furthermore, the education level of those individuals using the site can be considered 'high' since many of those frequenting the organizations various sites either attend or are employed by the institution. It cannot be assumed that such individuals possess a strong technical background since the vast majority of the websites hosted from the server are not technical in nature. The remaining two organizations under analysis, 'org2' and 'org3', are private companies who offer web-based business solutions through their websites. As such, the people who frequent these sites can be assumed to have a generally high knowledge of I.T. technologies and products; and can be considered as "advanced Internet users", who should be aware of issues such as their privacy and its implications when browsing the Internet.

Since P3P policies are XML files that require a significant degree of technical knowledge to interpret, it would seem reasonable to expect that users would choose to employ user agents to collect and analyze P3P policies. Currently, users have the choice of using a range of P3P agents or proxies:

- The default P3P agent installed in IE 6.0 and IE 7.0 which can provide a human readable form of the P3P policy when requested [12].
- AT&T Privacy Bird which is a plug-in to IE 6.0 [13]
- JRC P3P Proxy which can act as a proxy for any web browser [14].
- Privacy Fox plug-in for the Firefox web browser [15].
- Use Netscape Navigator 7.0 which can provide a human readable version of the P3P policy using the built-in 'Privacy Policy Viewer' [16].

Since we desire to analyze the number of unique individuals requesting P3P files, IP addresses will be used to approximate individual website patrons. While multiple users may make requests upon servers from the same IP address, no other feasible approach exists for identifying individual users through the information contained in server access logs.

5.4 Survey Results

Table 5-1 presents the results of the analysis from the server logs and indicates that P3P policies are almost never requested! The only server log that recorded any P3P requests is from Org1 and of the three recorded P3P requests, two were made from the same patron indicating that almost no one ever tests for P3P policies. This is in stark contrast with the active usage of other online signaling mechanisms such as E-Bays reputation mechanism [11] and even the observed rate that users reference human readable privacy policies [17]!

Hence, we must conclude that the first precondition has not been met, and that P3P protocol cannot be considered and effective signaling mechanism since Internet users lack the information required to discriminate between peach and lemon investors. The result will be a marketplace dominated by 'lemons' since they will be able to reap the greatest rewards through the misuse of their patrons information. Our model also predicts that no rational organization would adopt the P3P protocol when so few patrons test for it, hence we theorize that P3P has been adopted by agents under the belief that principles

Table 5-1: Analysis of the Number of Users Testing for P3P Policies

Organization	Time Period	Dates	Average number of requests per day	Average number of unique users per day	Total number of P3P requests	Total number of unique users requesting P3P documents
Org1	1-month	August 2006	409594	28854	3	2
Org2	15-months	June 2005 – October 2006	1598	51	0	0
Org3	34 days	Sept 3 2006 - Oct 6, 2006	4057	53	0	0

would begin testing shortly. Clearly, this belief has not born fruition and hence, our model predicts that:

- P3P adoption will remain stagnant, except where other external stimuli exist. A reduction in P3P adoption appears unlikely since the P3P signal has virtually no recurring costs.
- Corrective maintenance on invalid P3P documents will not be undertaken
- Little or no perfective maintenance will be undertaken on P3P policies because agents will have no motivation to increase the 'quality' of the signal they display.

Previous research [18], has found evidence supporting all three predictions of our model in marketplaces where laws do not explicitly require websites to adopt machine readable privacy policies. Reay *et al.* [18] found that valid P3P adoption by non-legislated websites not only stagnated, but actually decreased slightly between February and November 2005. Furthermore, Reay *et al.* [18] found that only 7 out of 609 non-legislated websites performed any corrective maintenance on their invalid policies over this time period! Similarly, perfective maintenance was only detected in 8 out of 514 non-legislated websites with valid full policies. This lack of perfective maintenance is especially surprising given that the vast majority of websites from the European Union, Canada, and Australia fail to meet the minimum requirements set forth in their jurisdictions national privacy legislation [7]. For instance, 65% of European websites retain information for an indeterminate time period in violation of Article 6.1.b of the European Data Directive [19]. Clearly room for improvement exists! These results suggest that Administrators of non-legislated websites appear to have little concern for the 'quality' of the P3P signal they provide.

The only deviation from these predictions occurred in U.S. government websites. This group of websites exhibited a large increase in P3P adoption occurred, had the lowest error rates for full P3P policies, and rarely adopted P3P compact policies without a corresponding full policy [18]. These findings suggest that the Administrators of these websites were concerned about signal quality and have taken extra care when adopting the P3P protocol. We can only conclude that these administrators believed that this additional effort would be rewarded in some form. We feel that these anomalous results are attributable to E-Governance Act [20] which requires all U.S. government websites to provide machine readable privacy policies. While similar requirements do not exist in other pieces of privacy protection legislation, many laws require organizations to inform principles as to how collected information will be used [19, 21]. Organizations may view the adoption of P3P protocol as a viable approach for persuading third parties that they are undertaking all reasonable means, sometimes referred to as 'best practice', to satisfy these requirements. Such actions, if sufficiently persuasive, may influence third parties into levying reduced sanctions when incidents occur.

5.5 Incorporating External Stimuli into privacy signaling models

In the previous section, it was noted that P3P adoption has in general stagnated. The single area where this picture breaks down is with regard to U.S. government web sites. It is hypothesized that these are experiencing a significant increase in adoption rates due to an external stimulus, specially the E-Governance act [20]. This external stimulus is beyond the initial model. Hence, in this section, we will extend the initial model to incorporate components, which will model such an external stimulus. While, the section specifically deals with the aforementioned situation, it is believed that the approach to extending the model is generic and can be applied to other external stimuli situations in an analogous fashion.

We propose to model the external stimuli created by third parties levying sanctions as a cost levied against agents (s) which creates agent disutility ($V(s)$). These costs may take the form of tangible monetary penalties or intangible damages to reputation. Agents may adopt signals in an attempt to persuade third parties that they respect their principles privacy thereby reducing or avoiding the sanctions third parties would otherwise levy where s_1 represents the reduced sanctions and s_2 represents the full sanctions. We model the act of persuading third parties using the mechanism used in Section 3 for the persuading of principles. Again, it is assumed that the third party will possess an "opinion" on the type of an agent given the signal it displays ($i(t)$) which will then be compared against a threshold value held by the third party (y) resulting in Equation 5-8:

Equation 5-8

$$s = \begin{cases} s_1 & \text{if} \quad i(t) \geq y \\ s_2 & \text{if} \quad i(t) < y \end{cases}$$

Incorporating this additional disutility into the previously developed model, results in Equation 5-9:

Equation 5-9

$$U_A = \begin{cases} U(p_2) - V(e) - V^R(t) - V(s_1) & \text{if} \quad q(t) \geq z \quad \text{and} \quad i(t) \geq y \quad \text{and} \quad \text{respects} \quad \text{privacy} \\ U(p_2) - V(e) - V^D(t) - V(s_1) & \text{if} \quad q(t) \geq z \quad \text{and} \quad i(t) \geq y \quad \text{and} \quad \text{disrespects} \quad \text{privacy} \\ U(p_2) - V(e) - V^R(t) - V(s_2) & \text{if} \quad q(t) \geq z \quad \text{and} \quad i(t) < y \quad \text{and} \quad \text{respects} \quad \text{privacy} \\ U(p_2) - V(e) - V^D(t) - V(s_2) & \text{if} \quad q(t) \geq z \quad \text{and} \quad i(t) < y \quad \text{and} \quad \text{disrespects} \quad \text{privacy} \\ U(p_1) - V(e) - V^R(t) - V(s_1) & \text{if} \quad q(t) < z \quad \text{and} \quad i(t) \geq y \quad \text{and} \quad \text{respects} \quad \text{privacy} \\ U(p_1) - V(e) - V^D(t) - V(s_1) & \text{if} \quad q(t) < z \quad \text{and} \quad i(t) \geq y \quad \text{and} \quad \text{disrespects} \quad \text{privacy} \\ U(p_1) - V(e) - V^R(t) - V(s_2) & \text{if} \quad q(t) < z \quad \text{and} \quad i(t) < y \quad \text{and} \quad \text{respects} \quad \text{privacy} \\ U(p_1) - V(e) - V^D(t) - V(s_2) & \text{if} \quad q(t) < z \quad \text{and} \quad i(t) < y \quad \text{and} \quad \text{disrespects} \quad \text{privacy} \end{cases}$$

This model of agent utility can be greatly simplified when the results of Section 5.4 are taken into account. Since users fail to test for the existence of the P3P signal, no rational agent will collect a payoff of p_2 since principles will not be able to identify privacy respecting websites. Thus, agents will only be provided with a payoff of p_1 resulting in Equation 5-10:

Equation 5-10

$$U_A = \begin{cases} U(p_1) - V(e) - V^R(t) - V(s_1) & if \quad i(t) \geq t \quad and \quad respects \quad privacy \\ U(p_1) - V(e) - V^R(t) - V(s_2) & if \quad i(t) < t \quad and \quad disrespects \quad privacy \\ U(p_1) - V(e) - V^D(t) - V(s_1) & if \quad i(t) \geq t \quad and \quad respects \quad privacy \\ U(p_1) - V(e) - V^D(t) - V(s_2) & if \quad i(t) < t \quad and \quad disrespects \quad privacy \end{cases}$$

Equation 5-10 can be further simplified if it is assumed that no rational agent will invest in a signal that fails to pass the requirements set forth by the third party leading to:

Equation 5-11

$$U_A = \begin{cases} U(p_1) - V(e) - V^R(t) - V(s_1) & if \quad i(t) \geq t \quad and \quad respects \quad privacy \\ U(p_1) - V(e) - V^D(t) - V(s_1) & if \quad i(t) \geq t \quad and \quad disrespects \quad privacy \\ U(p_1) - V(e) - V(s_2) & otherwise \end{cases}$$

If we assume that the signal is able to generate a separating equilibrium, which only occurs if $V^D(t) - V(s_1) > V(s_2)$, then Equation 5-11 can be further simplified resulting in:

Equation 5-12

$$U_A = \begin{cases} U(p_1) - V(e) - V^R(t) - V(s_1) & if \quad i(t) \geq t \quad and \quad privacy \quad respecting \\ U(p_1) - V(e) - V(s_2) & otherwise \end{cases}$$

In this model, only third parties will be able to reliably categorize agents into various types. From a principles perspective, a pooling equilibrium will still exist and the asymmetric information problem will remain. Privacy conscious principles will only take part in this market if they believe that all agents are privacy respecting. Since these beliefs are grounded in actual experiences, this requires most, if not all, agents to respect their patron's privacy. Such an environment will only develop if the sanctions levied by third parties are large enough to ensure that it is not in any agent's best interest to disrespect their principles privacy. However, the results from previous research [7] indicate widespread poor adherence to national regulations suggesting that national regulators have not been effective at applying sanctions for objectionable conduct. Furthermore, improvements in enforcement appear unlikely given the difficulty in harmonizing existing legal frameworks [22] and the variance in privacy concerns across cultures [7].

5.6 Conclusions

In this paper, we have investigated the effectiveness of the P3P protocol as a marketplace signal and have found that there exists no evidence suggesting that P3P has been effective in this role. This finding has raised an interesting question; if website patrons are not requesting these documents, why are websites providing them? Using results observed in previous research [18], we hypothesized that websites may be enacting P3P to persuade third parties that they are attempting to enact industry 'best practices' in the hope of limiting the magnitude of any sanctions levied against them if subsequent privacy disclosures occur. However, we have also found [7] in previous work that the vast majority of retrieved P3P policies appear to fail the minimum privacy standards set for by their respective legal jurisdictions. Furthermore, we have identified [18] that little if any maintenance is being undertaken on these policies, suggesting that either the policies are not being actively requested, or those who are requesting the policies are not investigating their contents. When one also takes into account the lack of significant increases in P3P adoption with the exception of members of the FirstGov list, it would appear that little motivation appears to exist for adopting the P3P protocol [18]. This problem highlights the need for greater analysis of what motivates end users into adopting privacy protection mechanisms.

Based upon these results, we can only conclude that P3P policies do not appear to play a significant role in the decision making processes of Internet users and many third parties. This evidence, and the lack of any observable motivational elements for continued P3P adoption, suggests that the P3P protocol has failed in its intended goal of providing Internet users with the information needed to protect their privacy.

5.7 Bibliography

[1] G. A. Akerlof, "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *Quarterly Journal Of Economics,* vol. 84, pp. 48-500, 1970.

[2] "On the Data Trail: How Detailed Information About You Gets Into The Hands of Organizations With Whom You Have No Relationship," The Canadian Internet Policy and Public Interest Clinic, Ottawa, Ontario, Canada April 2006.

[3] S. Lichtenstein, P. M. C. Swatman, and K. Babu, "Adding Value to Online Privacy for Consumers: Remedying Deficiencies in Online Privacy Policies with an Holistic Approach," in *36th Hawaii International Conference on System Sciences,* Hawaii, 2003.

[4] B. Fogg, J. Marshall, O. Laraki, A. Osipovich, C. Varma, N. Fang, J. Paul, A. Rangnekar, J. Shon, P. Swani, and M. Treinen, "What Makes Websites Credible? A Report on a Large Quantitative Study," in *SIGCHI 2001*, 2001.

[5] A. M. Spence, "Job Market Signaling," *Quarterly Journal Of Economics,* vol. 87, pp. 355-374, 1973.

[6] M. S. Ackerman, L. F. Cranor, and J. Reagle, "Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences," in *1st ACM Conference on*

Electronic Commerce, Denver, Colorado, 1999, pp. 1-8.

[7] I. Reay, S. Dick, and J. Miller, "A Large-Scale Empirical Study of Online Privacy Policies: Stated Actions vs. Legal Obligations," *ACM Transactions on the Web,* 2007, Under Review.

[8] M. Rothschild and J. E. Stiglitz, "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal Of Economics,* vol. 90, 1976.

[9] I. Molho, *The Economics of Information. Lying and Cheating in Markets and Organizations*, 1 ed. Malden, Massachusetts: Blackwell Publishing, 1997.

[10] M. Melnik and J. Alm, "Does a Seller's eCommerce Reputation Matter? Evidence from eBay Auctions," *Journal of Industrial Economics,* vol. 50, pp. 337-349, September 2002.

[11] P. Resnick and R. Zeckhauser, "Trust Among Strangers in Internet Transactions: Emperical Analysis of eBay's Reputation System," University of Michigan, 2001.

[12] Microsoft, "Microsoft Internet explorer (Pre-Release Version 7.0) Privacy Statement." vol. 2006: Microsoft, 2005.

[13] "Privacy Bird." vol. 2006.

[14] "JRC P3P Resource Center." vol. 2006: Joint Research Center.

[15] F. Arshad and S. Sheng, "Privacy Fox." vol. 2006, 2006.

[16] C. Corre, "Netscape 7.0 Preview Release 1." vol. 2006: Netscape Communications Corp, 2002.

[17] C. Jensen, C. Potts, and C. Jensen, "Privacy Practices of Internet Users: Self-Reports Versus Observed Behavior," *International Journal Human Computer*

Studies, vol. 63, pp. 203-227, 2005.

[18] I. Reay, P. Beatty, S. Dick, and J. Miller, "A Survey and Analysis of the P3P Protocol's Agents, Adoption, Maintenance, and Future," *Submitted for Review to IEEE Transactions on Dependable and Secure Computing,* 2006.

[19] "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data," in *23.11.1995*, 1995.

[20] "US E-Government Act of 2002, Public Law 107-347-DEC. 17 2002," 2002.

[21] "Personal Information Protection and Electronic Documents Act," Second ed, 2000.

[22] "First Report on the Implementation of the Data Protection Directive (95/24/EC)," Commission of the European Communities, 2003.

6 General Conclusion and Discussion

In order to enhance their competitive advantage, organizations have begun to take advantage of novel methods of information collection, aggregation, and usage. These actions often benefit both the organization and their customers by subsidizing the cost of goods and reducing advertising costs. Unfortunately, incidents of information misuse do occur, resulting in harm to customers. For instance, collected information may be used for: identity theft [1], user profiling [2], fraud [1], spam, and direct marketing [3, 4]. As these problems become more common place and their costs continue to rise [1], customer awareness will naturally increase, resulting in heightened online safety concerns [5-10]. Academia has responded to this growing public concern with a vibrant research community spread over multiple disciplines including engineering, business, economics, psychology, sociology, and law. The interdisciplinary nature of this topic has both helped and hindered academic analysis. The wealth of information generated from various standpoints fosters lively debate on this crucial topic, resulting in many novel proposals. Many of these proposals are limited to a single perspective though, resulting in ineffective solutions.

This thesis does not attempt to provide an exhaustive overview of this diverse topic, nor does it attempt to provide a comprehensive solution to the problem of online information security and privacy. We feel that neither are currently feasible since insufficient interdisciplinary research exists to support such actions. Instead, we attempt to fill a portion of this void with a series of interdisciplinary surveys using the P3P protocol as a vehicle for exploration

The use of the P3P protocol as a vehicle for exploration affords many unique benefits.

For instance, P3P allows researchers to undertake large scale, automated analyses of website privacy practices that would be infeasible to accomplish through the analysis of human readable privacy policies [11]. Furthermore, since P3P is an internationally recognized protocol, its usage facilitates the analysis of online privacy topics across national and cultural boundaries.

As previously mentioned, the body of this thesis was composed of four distinct studies each targeting a unique online privacy issue using the P3P protocol. The first study analyzed the adoption and maintenance of the P3P protocol by organizations. Results indicate that P3P adoption is limited and stagnate with the exception of U.S. government websites. We attributed this increase in U.S. government adoption to the enacting of the E-Governance Act [12] by the U.S. Federal government. In-fact, a slight decrease in P3P adoption was identified for non-legislated websites suggesting that further increases in P3P adoption appear unlikely. The most startling results in this study were the extremely high error rates and low P3P document maintenance. Furthermore, many organizations violate the P3P protocol by adopting P3P compact policies without full policies in violation of the P3P protocol. This finding suggests that many organizations adopted the P3P protocol in response to Internet Explorer's privacy protection utility which by default, blocks third party cookies not accompanied by a P3P compact policy. These findings paint a very bleak picture for the future of the P3P protocol as a viable privacy protecting technology since organizations appear to lack the motivation to adopt and maintain the protocol.

The error rates identified in this study raise an interesting question; if these documents contain so many structural errors, are they also filled with semantic errors? To our knowledge, no research exists which answers this question. Given that many future privacy enhancing technologies will likely attempt to reduce information asymmetry by sharing information in various formats, research in this area is urgently required.

The second study analyzed P3P adoption across cultural and national boundaries. Intercultural analysis of privacy enhancing technologies is crucial since cultures vary significantly in their perceptions of privacy [13] and little research currently exists investigating the topic [14, 15]. The results of this survey indicate that low-context cultures were far more likely to adopt the P3P protocol than their high-context counterparts and empirically support the work of Hsu and Kuo [14]. Furthermore, the results of this study also indicate that a cultures increased concern for information privacy is not strongly correlated with increases in P3P adoption suggesting that additional adoption factors exist for the P3P protocol. These results highlight the need for further research into how different cultural biases affect the adoption of other privacy enhancing technologies. The importance of such information is further heightened since our results also indicate that many of the assumptions held by individuals are likely incorrect. These issues must be addressed if privacy enhancing technologies are to experience widespread international adoption. Finally, this study also identified that the currently enacted national legislation appears to have had little or no harmonizing influence on the actions of organizations. This result is surprising given that the European Union adopted the Data Directive with the explicit goal of harmonizing the privacy practices of organizations from member nations [16]. This result also raised an interesting question; If legislation is not harmonizing the actions of organizations, is it having any noticeable effect? The third study was undertaken in an attempt to answer this question.

The results of the third study indicated that the stated actions of many organizations fail to meet the requirements set forth in the privacy legislation of their respective jurisdictions. Furthermore, the Safe Harbor program [17], appears to have had no detectable influence upon the actions of websites. Based upon these results, one can only conclude that current legislated attempts at protecting privacy have been met with mixed results at best. These results also bring into question whether APEC's current privacy harmonization proposal [18] will be effective given the significant cultural, economic, and legal differences which exist between member nations which are far greater than those between European Union nations [19, 20].

The final study investigated whether Internet users actively request P3P documents from websites. The observed lack of requests was very surprising given the number of websites that have adopted the protocol. In order to explain these results, we hypothesized that P3P adoption is driven not by end users, but by the actions of third parties such as government regulators. When the results of the first study are taken into account, we can only conclude that the P3P protocol can not be considered an effective privacy protecting technology. Future research is needed to identify the reasons for this failure.

This research has highlighted several other potential avenues for future research that leverage the P3P protocols strengths. For instance, it has been shown that P3P can be used as an effective tool for large scale automated analysis of organizational privacy sensitive actions. With some modifications, a P3P like tool could greatly aid national governments in investigating privacy legislation adherence. The modifications must be carefully considered though since such a tool needs to possess sufficient descriptive power and yet remain flexible so that the tool can be adapted to evolving legislation.

Potential solutions to these problems may lie with semantic web technologies [21].

The consideration of both cultural and legal concepts in this thesis provides a unique perspective on the topic of privacy which is usually lacking in current research into privacy enhancing technologies. Cultural analysis allows for a macroscopic perspective of how societal beliefs affect technology adoption. We identified that websites from low context nations were significantly more likely to adopt P3P than their high context counterparts suggesting that future privacy enhancing technologies need to be resistant against such a cultural divide. In contrast, a legal analysis allows for a comparatively microscopic analysis of societal influences on technology. Laws are in essence, a codification of a subset of all the values a culture holds. Hence, legal analysis allows for a narrower, more rigorously defined perspective on the issues in question. Given that our results suggest that many of these laws appear to be at least partially ignored, an interesting question is raised; Which has a greater influence on an organizations privacy sensitive actions, culture or law? We currently lack the resources to answer this question, but future research needs to consider this question. An answer would benefit both the creators of public policy, as well as information asymmetry minimization technologies.

6.1 Bibliography

[1] Synovate, "Federal Trade Commission - Identity Theft Survey Report," September 2003.

[2] H. Black, "On-Line Data Brokers." vol. 2006: Canadian Privacy Commissioner, 2005.

[3] H. Kargupta, A. Joshi, K. Sivakumar, and Y. Yesha, *Data Mining: Next Generation Challenges and Future Directions*. Menlo Park, California: MIT

Press, 2004.

[4] C. R. Taylor, G. R. Franke, and M. L. Maynard, "Attitudes Toward Direct Marketing and Its Regulation: A Comparison of the United States and Japan," *Journal of Public Policy and Marketing,* vol. 19, pp. 228-237, 2000.

[5] S. Spiekermann, J. Grossklags, and B. Berendt, "E-Privacy in 2nd Generation E-Commerce: Privacy Preferences Versus Actual Behavior," in *3rd ACM conference on Electronic Commerce*, 2001.

[6] P. Kumaraguru and L. Cranor, "Privacy in India: Attitudes and Awareness," in *2005 Workshop on Privacy Enhancing Technologies*, Dubrovnik, Croatia, 2005.

[7] C. Jensen, C. Potts, and C. Jensen, "Privacy Practices of Internet Users: Self-Reports Versus Observed Behavior," *International Journal Human Computer Studies,* vol. 63, pp. 203-227, 2005.

[8] R. K. Chellappa and R. G. Sin, "Personalization versus Privacy: An Emperical Examination of the Online Consumer's Dilemma," *Information Technology and Management,* vol. 6, pp. 181-202, 2005.

[9] H. Nissenbaum, "Protecting Privacy in an Information Age: The Problems of Privacy in Public," *Law and Philosophy,* vol. 17, pp. 559-596, 1998.

[10] M. S. Ackerman, L. F. Cranor, and J. Reagle, "Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences," in *1st ACM Conference on Electronic Commerce*, Denver, Colorado, 1999, pp. 1-8.

[11] J. B. Earp, A. I. Antón, L. Aiman-Smith, and W. H. Stufflebeam, "Examining Internet Privacy Policies Within the Context of User Privacy Values," *IEEE Transactions on Engineering Management,* vol. 52, pp. 227-237, May 2005.

[12] "US E-Government Act of 2002, Public Law 107-347-DEC. 17 2002," 2002.

[13] R. Capurro, "Privacy. An Intercultural Perspective," *Ethics and Information Technology,* vol. 7, pp. 37-47, 2005.

[14] M.-H. Hsu and F.-Y. Kuo, "The Effect of Organization-Based Self-Esteem and Deindividualism in Protecting Personal Information Privacy," *Journal of Business Ethics,* vol. 42, pp. 305-320, February 2003.

[15] N. Zakaria, J. M. Stanton, and S. T. M. Sarkar-Barney, "Designing and Implementing Culturally-Sensitive IT Applications," *Information Technology and People,* vol. 16, 2003.

[16] "First Report on the Implementation of the Data Protection Directive (95/24/EC)," Commission of the European Communities, 2003.

[17] "Safe Harbor Program." vol. 2006.

[18] "APEC Privacy Framework," Asia-Pacific Economic Cooperation, Santiago, Chile 2004/AMM/014rev1, 2004.

[19] E. T. Hall, *Beyond Culture.* New York: Doubleday, 1989.

[20] G. Hofstede, "Cultural Dimensions." vol. 2006, 2006.

[21] R. Benjamins, *Law and the Semantic Web: legal ontologies, methodologies, legal information retrieval, and applications.* New York: Springer, 2005.

7 Appendix 1. Latent Class Analysis

Latent Class analysis is a form of Latent Structure analysis. Through latent class analysis techniques, Observed nominal and ordinal responses to various questions are mapped onto a latent variable. Each value of the latent variable represents a class of responses that fall upon a continuum of possible responses.

To understand the process behind Latent Class analysis, we present a modification of an example originally given by McCutcheon [1]. The example involves the simplest latent class problem that can exist; identifying a latent variable that explains the relationship between two observed variables (Table 7-1).

In this example Table 7-1 represents the results of a fictitious survey which attempted to determine if patients given Drug A had an increased likelihood of survival. Through an application of Fishers Exact Test, it can easily be determined that a statistically significant difference exists in the survival rate of people who are given the drug and those that are not ($p < 0.0053$). This statistically significant difference indicates that the variables 'Survived' and 'Given Drug A' are not independent of each other. In other words, there is a relationship between survivability and whether the patient was given a drug.

Table 7-1: Contingency Table for Two Item Example

		Given Drug A	
		Yes	No
Survived	Yes	95	55
	No	70	80

Suppose that there was an unobserved factor that the survey did not take into account (ex. were patients taking Drug B or not). This factor, while not directly observed, may be able to explain the relationship between patient survivability and Drug A. If the relationship between Survivability and Drug A is explained when Drug B is taken into account, then we say Drug B is a latent factor.

If we assume that Drug B was also taken into account, the two-way comparison becomes a three-way comparison which is represented by the Cross tabulation table depicted in Table 7-2. The numbers in Table 7-2 are a hypothetical distribution which could occur if Drug B was also taken into account in Table 7-1. It can be easily verified that Table 7-1 can be generated from Table 7-2 by removing the variable representing Drug B and merging the relevant columns. If each level of B is analyzed individually (Table 7-3 and Table 7-4), it can be easily verified that the statistically significant relationship between survivability and Drug A no longer exists (p = 1.000 in both cases). Thus, the latent variable Drug B, explains the relationship between the two variables. If a different hypothetical distribution was used when taking into account Drug B such as the distribution observed in Table 7-5, then the previously observed relationship may still exist since a statistically significant difference still exists for Table 7-6 ($p < 0.0484$). In this instance, Drug B must be rejected as a latent factor since it cannot explain the observed relationship.

Table 7-2: Cross Tabulation table for 3 Variable Analysis

		Given Drug B		Not Given Drug B	
		Given Drug A	Not given drug A	Given Drug A	Not given drug A
Survived	Yes	80	20	15	35
	No	40	10	30	70

Table 7-3: Cross Tabulation Table Given that the Patients on Drug B

		Given Drug A	
		Yes	No
Survived	Yes	80	20
	No	40	10

Table 7-4: Cross Tabulation Table Given that the Patients were not on Drug B

		Given Drug A	
		Yes	No
Survived	Yes	15	35
	No	30	70

Table 7-5: Another Hypothetical Distribution for taking into account being given Drug B

		Given Drug B		Not Given Drug B	
		Given Drug A	Not given drug A	Given Drug A	Not given drug A
Survived	Yes	48	27	47	28
	No	35	40	35	40

Table 7-6: Cross Tabulation Table Given that the Patients were on Drug B

		Given Drug A	
		Yes	No
Survived	Yes	48	27
	No	35	40

Table 7-7: Cross Tabulation Table Given that the Patients were not on Drug B

		Given Drug A	
		Yes	No
Survived	Yes	47	28
	No	35	40

When a latent factor is able to explain the observed relationship between two or more variables, the variables are said to be locally independent given the latent variable. Local independence is the fundamental theory underpinning Latent class models. If the latent variable is able to generate locally independent populations within its various latent classes, then a suitable latent structure for the observed data set has been identified.

7.1 Axiom of Local Independence

All latent structure models attempt to locate areas clusters of responses which are locally independent. The Axiom of Local Independence is the foundation of these models whereby all models assume that the observed variables are independent of each other given a latent variable. Harper [2] explains that this assumption states in effect that

> "the items together constitute a "pure" test – a test which measures one attribute and is free from other contaminating or extraneous factors" pg. 53.

Since such a pure test cannot be employed in practice, the assumption rarely holds. However, if tests are close to meeting this assumption, little difference will exist between the model and actual observations. If this assumption is badly met, then models based upon this assumption will map poorly onto observations. This variability in the ability of the model of fit observed data provides a mechanism to test individual models for sufficient fit. However, before such a test can be developed, a method must be developed to apply the axiom to the observed dataset.

The Axiom of Local Independence is described as:

Equation 7-1

$$\pi_i = \sum_\alpha v^\alpha \lambda_i^\alpha$$

Equation 7-2

$$\pi_{ij} = \sum_\alpha v^\alpha \lambda_i^\alpha \lambda_j^\alpha \qquad i \neq j$$

Equation 7-3

$$\pi_{ijk} = \sum_\alpha v^\alpha \lambda_i^\alpha \lambda_j^\alpha \lambda_k^\alpha \qquad i \neq j, \quad i \neq k, \quad j \neq k$$

Where π_i is the proportion of observations at a certain level for item *i*. v^{α} is the proportion of people in class α, and λ_j^{α} is the probability that an observation in class α will be at a certain level for item *i*. By the axiom, the probability that an observation in class α will be at a certain level for items *i* and *j* is given by the product of their probabilities $\lambda_i^{\alpha}\lambda_j^{\alpha}$ (Equation 7-2). Similar logic applies when 3 items are considered (Equation 7-3). Thus, if some method can be developed for determining these probabilities, the Axiom of Local Independence can predict the proportion of observations that should be expected based upon variable values, given a certain class and locally independent populations.

7.2 Latent Class Model

The Latent Class Model is an instance of the general Axiom of Local Independence. For the previous example, the previous example would be locally independent if:

Equation 7-4: Latent Class Model for Above Example

$$\pi_{ijt}^{SAB} = \pi_{it}^{\bar{S}B} \times \pi_{it}^{\bar{A}S} \times \pi_{t}^{S}$$

Where 'S' represents the variable indicating whether people survived and 'i' represents the dichotomous response. 'A' represents whether people were given Drug A and 'j' represents the dichotomous response. Finally, 'B' represents the latent variable which indicates whether people were given Drug B and 't' represents the dichotomous response. π_{ijt}^{SAB} represents the probability that a randomly selected response pattern will be located in the i, j, t cell of Table 7-2 or Table 7-5. $\pi_{it}^{\bar{S}B}$ is the conditional probability that the randomly selected response pattern has value 'i' given that it belongs to class 't'. Finally,

π_t^B is the latent class probability and indicates the probability that a randomly selected response pattern belongs to the latent class ‘t’.

A general form of the Latent Class Model is defined as:

Equation 7-5: General Latent Class Model

$$\pi^{AB...EX}_{ij...mt} = \pi^{\bar{A}X}_{it} \times \pi^{\bar{B}X}_{it} \times ... \pi^{\bar{E}X}_{mt} \times \pi^{X}_{t}$$

Where the expected value of the cell in the above cross tabulation tables is given by:

Equation 7-6: Expected Cell Value assuming Local Independence

$$\hat{F}_{ij..mt} = \pi^{AB..EX}_{ij..mt} \times SampleSize$$

When Equation 7-5 holds, the axiom of local independence predicts that the expected cell value (Equation 7-6) will precisely match the observed cell value in the cross tabulation table. As the assumption of local independence becomes ill founded, the result of Equation 7-6 will begin to deviate from the observed value in the Cross Tabulation table.

An example is provided below where we calculate the expected frequency that a patient survived, was given drug A, and given drug B for both Table 7-2 (Figure 7-1) and Table 7-5 (Figure 7-2). In the two Figures, π_{11}^{SB} represents the conditional probability that a patient survived if they were given Drug B. π_{11}^{AB} represents the conditional probability that a patient was given Drug A if they were given Drug B. π_1^B represents the latent class probability that a patient was given Drug B. π_{111}^{SAB} represents the probability that a patient survived, was given Drug A, and given Drug B. Finally, F_{111}^{SAB} represents the expected cell frequency that should be observed in the cross tabulation table.

It can be easily observed that when local independence exists (Figure 7-1), the latent class model is able to precisely predict the cell value in Table 7-2 (80). However, when local independence is lacking (Figure 7-2), the prediction of the latent class model (42) strays from the observed cell value in Table 7-5 (48).

$$\pi_{11}^{SB} = \frac{80+20}{150} = 0.6667$$

$$\pi_{11}^{AB} = \frac{80+40}{150} = 0.8000$$

$$\pi_{1}^{B} = \frac{150}{300} = 0.5000$$

$$\pi_{111}^{SAB} = 0.6667 \times 0.8000 \times 0.5000 = 0.2668$$

$$F_{111}^{SAB} = 0.2668 \times 300 = 80.00$$

Figure 7-1: Calculation of the Expected Frequency for the Cell Indicating the Number of Patients Who Survived, Were Given Drug A, and Given Drug B from Table 7-2

$$\pi_{11}^{SB} = \frac{48+27}{150} = 0.5000$$

$$\pi_{11}^{AB} = \frac{48+35}{150} = 0.5533$$

$$\pi_{1}^{B} = \frac{150}{300} = 0.5000$$

$$\pi_{111}^{SAB} = 0.5000 \times 0.5533 \times 0.5000 = 0.1383$$

$$F_{111}^{SAB} = 0.1383 \times 300 = 42$$

Figure 7-2: Calculation of the Expected Frequency for the Cell Indicating the Number of Patients Who Survived, Were Given Drug A, and Given Drug B from Table 7-5

7.3 Latent Class Probabilities

From the preceding discussion, it can be observed that the latent class probabilities play a central role in the latent class model. In addition to being used in the latent class model, these probabilities also provide information regarding the number of latent classes and their relative sizes [1]. For instance, if a one latent class model is able to accurately represent the population under analysis, then the only conclusion that can be made is that the observed variables are already independent of each other. Similarly, if a single latent class is far greater in magnitude than the other classes, a Researcher may conclude that the smaller classes represent outlier effects.

7.4 Conditional Probabilities

The Conditional Probabilities are comparable to factor loadings in traditional factor analysis. These probabilities indicate the probability that a randomly selected observation will have a particular variable at a certain value given the observation belongs to a certain latent class. These probabilities allow researchers to identify unique attributes of particular latent classes.

7.5 Expectation Maximization (EM) Model

For the latent class models to be of value, a method is required to determine the conditional and latent class probabilities. Two popular approaches exist for the estimation of these probabilities. The first is the Newton-Raphson approach. This approach, while computationally intensive, allows for restraints on the conditional and latent class probabilities and finds asymptotic covariances [3]. The second approach is Goodman's Maximum Likelihood Estimation procedure [4] which is a special case of the more general 'EM algorithm' [5]. This method, utilizes an iterative approach whereby an expectation 'E' is first developed which is then maximized 'M' thus leading to the 'EM' name for the general approach. The use of Goodman's method ensures convergence to at least a local minimum and guarantees that all probability estimates fall on the interval 0-1 [3] which the Newton-Raphson method cannot guarantee. When undertaking exploratory class analysis, no restrictions on conditional and latent class probabilities are required but convergence and assurances of the interval estimates fall under are very desirable. For this reason, we choose to employ Goodman's Maximum Likelihood Estimation procedure.

The 'EM algorithm' operates by making expectations (generating a model) based upon the observed data, and then attempting to maximize a test statistic. In Goodmans method, the expectation phase (of the general 'EM' algorithm) involves the researcher proposing initial latent class and conditional probabilities for the General latent class model (Equation 7-5). In exploratory latent class analysis, these initial values are often randomly generated since researchers are not required to have a theoretical latent structure upon which to derive their initial estimates.

The maximization phase, involves improving the models estimates through an iterative procedure which is explained below. This discussion will use the previous example where two observed variables are used to estimate the latent variable. Thus, the general latent class model for the problem is given by Equation 7-7. The bar above the pi symbol indicates that the probabilities are estimates. For readers interested in the Formal Derivation of the below procedure, see Goodman [4].

Equation 7-7:

$$\bar{\pi}^{SAB}_{ijt} = \bar{\pi}^{\bar{S}B}_{it} \times \bar{\pi}^{\bar{A}B}_{it} \times \bar{\pi}^{B}_{t}$$

Since we are undertaking exploratory analysis, the initial values for the conditional ($\bar{\pi}^{SB}_{it}$, $\bar{\pi}^{AB}_{it}$) and latent class $\bar{\pi}^{B}_{t}$ probabilities will be generated by a pseudo-random number generator.

The first Step of the maximization procedure involves summating the conditional probabilities of all latent classes for a particular cell (Equation 7-8). The results of Equation 7-8 create a prediction of what proportion of observations should occur in each cell.

Equation 7-8

$$\bar{\pi}_{ij} = \sum_{t} \bar{\pi}^{SAB}_{ijt}$$

These values can be used to generate an estimation of the probability that an observation at levels i, j for variables S, A will be at level t (Equation 7-9).

Equation 7-9

$$\bar{\pi}^{SA\bar{B}}_{ijt} = \frac{\bar{\pi}^{SAB}_{ijt}}{\bar{\pi}_{ij}}$$

Using the observed cell values, new latent class probabilities can be produced by Equation 7-10 where p_{ij} represents the proportion of observations in the i,j cell (Table 7-1)

Equation 7-10

$$\bar{\pi}^{B}_{t} = \sum_{ij} p_{ij} \bar{\pi}^{SA\bar{B}}_{ijt}$$

These latent class probabilities can then be used to generate new conditional probabilities:

Equation 7-11

$$\bar{\pi}^{\bar{S}B}_{it} = \frac{\sum_{j} p_{ij} \bar{\pi}^{SA\bar{B}}_{ijt}}{\bar{\pi}^{B}_{t}}$$

Equation 7-12

$$\bar{\pi}^{\bar{A}B}_{jt} = \frac{\sum_{i} p_{ij} \bar{\pi}^{SA\bar{B}}_{ijt}}{\bar{\pi}^{B}_{t}}$$

Once the conditional probabilities are generated, the resulting model is compared to the previous model using the chi-square log-likelihood test. If the difference between the two models is less than a specified stopping criterion, the process is stopped, otherwise the process loops back to Equation 7-7 and the process repeats. Since such an algorithm could conceivably look a significant number of times, a maximum number of iterations is usually specified by the researcher.

7.6 Exploratory Latent Class Analysis

Exploratory Latent Class Analysis employs the above methods on data sets where the researcher has little or no expectations regarding the structure of the latent model. Thus, when undertaking ELCA, the researcher usually uses random values to create their initial expectations. Once the initial probabilities are created, the process proceeds as described. Since this model does not require the researcher to have any expectations regarding the structure of the latent model, methods are required to identify models exhibiting sufficient fit as well as choosing a single model out of a large number of well fitting models.

The testing of model fit is done through a comparison of the observed values, and the model predictions which can be calculated by Equation 7-13. This comparison is usually accomplished by applying the likelihood-ratio chi-square test however, depending upon the type of sample, other tests such as Pearson's chi-square or Fishers exact test could also be appropriate. If a statistically significant deviation is detected between the Model predictions and the observed cell values, then the model must be rejected.

Equation 7-13

$$F_{ij} = \overline{\pi}_{ij} * SampleSize$$

The process of identifying which model to use in Exploratory Latent class analysis involves a sequential elimination of simple models in favor of more complex models (additional latent classes). Thus, the first step in ELCA determines whether a one latent class model provides sufficient fit. If this model fits, then the data set already exhibits local independence forcing the researcher to conclude that the variables are independent. If a one latent class model does not provide sufficient fit, then the researcher proceeds to test a 2 latent class model. If a two latent class model does not fit, the researcher then tests a 3 latent class model. This process continues until a model is found that does not deviate from the observed cell values beyond what random chance would allow for.

7.7 Bibliography

[1] A. L. McCutcheon, *Latent Class Analysis*. Newbury Park: Sage Publications, Inc, 1987.

[2] D. Harper, "Local Dependence Latent Structure Models," *Psychometrika,* vol. 37, pp. 53-60, 1972.

[3] A. Mooijaart and P. G. M. v. d. Heijden, "The EM Algorithm for Latent Class Analysis With Equality Constraints," *Psychometrika,* vol. 57, pp. 261-269, 1992.

[4] L. A. Goodman, "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifable Models," *Biometrika,* vol. 61, pp. 215-231, August 1974.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from

Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological),* vol. 39, pp. 1-38, 1977.