APPROVAL SHEET

Title of Dissertation: Risk Analysis of the Discoverability of Personal Data Used for Primary and Secondary Authentication

Name of Candidate: Kirsten Esther Richards

Doctor of Philosophy, Information Systems 2017

Dissertation and Abstract Approved: _____

Anthony F. Norcio, Ph.D.

Professor

Information Systems

Date Approved: _____

ABSTRACT

Title of Document:	RISK ANALYSIS OF THE DISCOVERABILITY OF PERSONAL DATA USED FOR PRIMARY AND SECONDARY AUTHENTICATION
	Kirsten Esther Richards, Ph.D. Information Systems 2017
Directed By:	Professor Anthony F. Norcio, Information Systems

Personal data are frequently leveraged to create passwords for password based authentication systems. Personal data are also used in secondary authentication systems, particularly those based around a question and answer format. The use of personal data in authenticators is believed to be driven, to some degree, by usability. The antinomic proposition of usable system authentication, an easily remembered and usable scheme for the proper user which is simultaneously unknown and unusable to any other entity, historically proves to be an elusive goal. While alternative propositions for authentication protocols are numerous, lacking in literature is foundational work directly relating potential authenticators with the discoverability of personal data online. This dissertation investigates the discoverability of personal data, particularly whether another human is able to purposefully find particular personal data commonly used in authentication protocols. Between fifty and sixty participants provide search results for specific personal data regarding four additional participants. The four participants

acted as a source for the personal data, consented to the web search and validated the accuracy of data supplied by the data seeking participants. Analyses of the results reveals consistent patterns in the personal data discovered. The results lay a foundation for the improvement of current authentication systems and provide a significant step in both methodology and recommendations to guide the development of alternatives with a goal towards the creation of usable, secure authentication systems. Furthermore, the results provide insight into the nature of privacy, user control of data and the availability of personal data on Web sources.

RISK ANALYSIS OF THE DISCOVERABILITY OF PERSONAL DATA USED FOR PRIMARY AND SECONDARY AUTHENTICATION

by

Kirsten Esther Richards

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, Baltimore County, in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy in
Information Systems

2017

Acknowledgements

Thank you to the wonderful faculty of UMBC's Information Systems Department, all of whom have contributed to my education in some way. A special thank you to my committee members, Dr. Joyram Chakraborty of Towson University, Dr. Richard Forno from UMBC's Department of Computer Science and Electrical Engineering, Dr. Wayne Lutters and Dr. Lina Zhou for generously giving of your expertise and time. Thank you to Dr. Aryya Gangopadhyay for the opportunity to participate in the GAANN fellowship. A special thank you to Dr. Anthony Norcio. The influence of your exemplary mentorship cannot be overestimated.

Thank you to Barb Morris, Ann Stavely, and Shannon Keegan for the crucial support you provide. You contributed greatly to my success.

Thank you to my husband, Tim Richards and my children, Anna, Abigail and Elizabeth, for your patience and support. Without you, this work would never have been finished. Finally, thanks to God, whose grace and provision made this work possible.

Contents

List of Tables

List of Abbreviations

HCC - Human Centered Computing

HCI - Human Computer Interaction

IS – Information Systems

SNS - Social Networking Sites

PCCP - Persuasive Cued Clickpoints

URL – Uniform Resource Locator

Chapter 1: Introduction

The design of feasible authentication mechanisms able to provide adequate security and privacy while maintaining usability remains an elusive goal for the Information Systems (IS) community. Considering usability and security in tandem creates an hitherto intractable problem requiring a broad spectrum of research specialties including Human Centered Computing (HCC) and Human Computer Interaction (HCI). While conceived and introduced frequently, various authentication schemes have not succeeded in wide-spread replacement of the password, despite superiority to standard passwords with regard to usability, security, or other factors considered individually (Bonneau, Herley, van Oorschot, & Stajano, 2012; Furnell & Zekri, 2006; Herley & Van Oorschot, 2012; Vander Veen, 2013). Alternative forms of authentication demonstrate various levels of promise with regard to usability or improving the reliability of identifying the user correctly and providing appropriate and accurate systems and data access. However, despite much dedicated effort, passwords remain the most common form of authentication and seem likely to remain in use for some time (Bonneau et al., 2012; Herley & Van Oorschot, 2012).

In addition to other security weaknesses, password usage is fraught with poor user behaviors that compromise security (Besnard & Arief, 2004). Failure to change passwords frequently, use of the same password on multiple devices, weak passwords with insufficient randomness or number of characters, and writing down or sharing passwords with others may compromise security (Anderson & Agarwal, 2010; C. Connelly, Archer, Yuan, & Guo, 2011; Keith, Shao, & Steinbart, 2007). These user behaviors are attributed to a wide variety of causes including user laziness, lack of information or training (Duggan, Johnson, & Grawemeyer, 2012; Vander Veen, 2013). Research also supports a variety of emotional, psychological, and cognitive explanations for user behaviors (Anderson &

Agarwal, 2010; C. Connelly et al., 2011; Gulenko, 2014). Despite persistent attempts to change user behavior through user education, policy requirements, clever mnemonic devices, and a wide variety of conflicting recommendations from the technical community, research routinely demonstrates that user behavior remains fundamentally problematic to the secure implementations of passwords (Anderson & Agarwal, 2010).

One user behavior that compromises the security of passwords is the frequent reliance upon personal knowledge for password creation. Information such as a birthdate or anniversary is often incorporated into a password (Brown, Bracken, Zoccoli, & Douglas, 2004). Additionally, many forms of secondary authentication rely upon personal data by design (Furnell, 2005). The combination of the use of personal data for primary and secondary authentication becomes potentially problematic when considering the immense volume and diversity of personal data available online (Acquisti & Gross, 2006). Some connections have been made between personal information known to acquaintances and secondary authentication (Schechter, Brush, & Egelman, 2009). However, no current studies directly address the connection between personal data available online and personal data used in secondary and primary authentication.

The particular problem of addressing password difficulties directly, as opposed to suggesting alternatives to passwords, is under-developed in research (Herley & Van Oorschot, 2012). This work begins to ameliorate the knowledge gap by improving an understanding the relationship between personal available information and the personal data commonly used in primary and secondary authentication. Furthermore, this work seeks to evaluate the ability of users to easily and accurately locate personal data and identity sources of data vulnerability to stranger attacks from public data sources.

1.1 The Problem of Password Based Authentication

The HCC community represents a long standing tradition of emphasizing a meaningful attention to user needs and capabilities and "freeing users to do their work" (Grudin, 2008). Studies highlight the movement away from design that assumes users are "stupid" or primarily responsible for problems (Grossman, 2009). Passwords represent a fundamental incongruity between technical systems requirements and the ability to support the needs and limitations of users, particularly in the context of daily life. Current authentication mechanisms require users to devote a high level of attention to security, rather than the multiplicity of tasks perceived as more important by the user. These tasks are perhaps an integral part of the continued difficulty of recruiting users to contribute to security through good password practices (Grawemeyer & Johnson, 2011; Pavlou, 2011). Human behaviors, potentiated by the continued use of password-driven authentication technology, will continue to generate security and privacy risks creating a weakest link in security (Herath & Rao, 2009; Sasse, Brostoff, & Weirich, 2001).

Passwords are a fundamentally flawed authentication mechanism from a human cognitive perspective. They fail to utilize human strengths to improve digital privacy and security while relying simultaneously on users to perform difficult tasks to maintain system and data integrity. This reality is well illustrated in the characteristics of a good password. Good password creation relies upon humans to perform cognitive tasks at which they are notoriously inept, thus betraying a poor interaction design from an HCI perspective. The cognitive difficulty of passwords is apparent in a description of a good password. Ideal passwords draw from a large character set, avoid common and dictionary words, are lengthy (Keith et al., 2007), combine upper and lowercase letters, use numbers and special

characters in the middle of the password (Brown et al., 2004). Humans are inherently limited in their channel capacity, both physically and cognitively (Baddeley, 1994). The randomness and length of the ideal password are juxtaposed to the strengths of human memory. Recollection relies upon patterns and relatively short sets or "chunks" of information for both learning and memory and the "magical number seven, plus or minus two" is a long standing perspective on human working memory limitations (Miller, 1956 p. 81). Recent work suggests that even shorter volumes of information sets are more accurate reflections of human memory (Cowan, 2015; Gignac, 2015; Miller, 1956). Additionally, as already discussed, many users are required to use multiple passwords further stretching their cognitive limitations. Advice from security expects, such as avoiding writing passwords down and using patterns in passwords, prevents humans from using standard memory aids relied upon in other contexts (Tam, Glassman, & Vandenwauver, 2010).

1.2 Problem of Personal Information Employed in Authentication

In light of the challenges passwords present to human cognition, it is not surprising that users resort to alternative resources to remember their passwords and retain access to systems. Passwords are seldom created to be "cryptic" – that is, difficult to decipher (Andrews, 2002; Brown et al., 2004). Instead, studies in human behavior in password creation and management reveal that actual usage seldom results in security conscious passwords. Brown et al. discuss Andrews' study of 1,200 individuals which found that only 10% of passwords were cryptic while the majority of passwords were designed for memorability (Andrews, 2002; Brown et al., 2004). According to Campbell, et. al policy initiatives do not significantly reduce the inclusion of personal information, such as names

and birthdates, in password creation (J. Campbell, Ma, & Kleeman, 2011). The use of personal information in password creation exposes the password to security threats through the use of informed guessing based on a combination of knowledge of common password patterns and garnered personal information of the account holder.

Additionally, Brown et al. assert that a third of the participants in their study forgot their password, some of those regularly (Brown et al., 2004). The need to retrieve passwords frequently and restore access to accounts exposes authentication to another personal data threat. Many secondary authentication mechanism are used to restore password access, one of the most popular being the correct answer to a question, often a question about personal data or guessable questions, for example "what is your favorite song?" (Reeder & Schechter, 2011). Thus, the primary method of authentication currently in use as well as the recovery mechanisms for failure remains heavily reliant on the use of personal data.

1.3 Problem of Available Personal Information

The availability of personal data and the contribution of personal data to account compromise are well-recognized problems (Dlamini, Eloff, & Eloff, 2009; Oravec, 2012; Pavlou, 2011; Schneier, 2010). Wide spread government and corporate use and publication of data as well as information compromised by hacking and other illegal activities contribute to the availability of personal data. An additional development of security compromise arises through the self-propagation of personal data made available widely on the internet (Schneier, 2010). Neither criminal skill, nor advanced technical knowledge, nor even rudimentary understanding of password cracking tools, is required to access this personal information which is believed to be attainable readily. As already discussed, some

of this personal data may be used in either primary or secondary authentication (Reeder & Schechter, 2011). This can compromise information secured through the use of personal data in authentication protocols.

Exposure may also occur through venues such as online social media, where an innocuous post regarding a personal event, such as where a user met their spouse, could conceivably allow for a correct guess using obtainable information (Reeder & Schechter, 2011). This information may then be used to access account information through guessing password combinations or answering secondary authentication security questions and resetting passwords or redirecting account information to accounts controlled by the hacker (Reeder & Schechter, 2011). However, the level of personal exposure and danger to non-celebrity individuals may vary widely. Numerous studies have explored information sharing and self-disclosure as well as comparisons between privacy attitudes and sharing and describe a wide breadth of information sharing behaviors (A. Acquisti & Gross, 2006; Beldad, de Jong, & Steehouder, 2011). Compounding the security problem presented by wide spread personal data sharing through online venues, is the failure of privacy controls on popular Social Networking Sites (SNS) to accurately reflect user intentions (Madejski, Johnson, & Bellovin, 2012). Copious personal information, for many Americans, is continuously collected and stored in many venues and may be compromised in a wide variety of situations (Il-Horn, Kai-Lung, Sang-Yong Tom, & Png, 2007).

Lacking in research is a clear relationship between personal data stored online in a variety of public or semi-public venues such as SNS and personal information used in authentication protocols. Furthermore, the ease of discovering personal data by other individuals and the Web locations posing the greatest practical threat are unknown. Thus,

while personal information online poses a credible threat to security accomplished through password protection and secondary authentication using personal data; the empirical extent of the threat is unknown.

1.4 Problem Summary

In summary, the majority of authentication approaches currently in use is accomplished through passwords. There are many other potentially viable authentication schemes which are discussed in detail in Chapter 2. None currently enjoys popular adoption in the majority of systems and several of them are also subject to personal data security flaws (Bonneau et al., 2012). While the continued use of passwords remains problematic in many ways, two significant aspects of the password problem are addressed here.

First, the difficulty of using passwords from a human cognitive perspective motivates users to select passwords for memorability rather than security. This human need encourages the reliance of users, in password creation and maintenance, on personal information. This problem extends beyond primary authentication to secondary authentication, which also relies frequently upon personal information to validate identity. The problem with personal data in secondary authentication often exists by design, not always by user choice.

Second, personal information for many individuals is believed to be widely available. The rise of social media, blogging and other forms of information sharing, in addition to corporate and government data collection provide an unprecedented venue to obtain the types of personal information frequently used in primary and secondary authentication. This combination of circumstances exposes users to security and privacy compromise through account password guessing based on information obtained from

publicly or widely available sources. Secondary authentication questions based on personal data also expose accounts to security compromise through the use of available personal data. However, the individual level of exposure and ease of discovering the specific information frequently used in password creation and secondary authentication has not been thoroughly explored.

1.5 Significance of the Study

The extensive use of passwords, often composed of personal data, combined with readily available personal data presents a security and privacy problem hazardous to both individuals and organizations. Breaches of security and privacy occur regularly and are costly (Belanger & Crossier, 2011; Gupta & Sharman, 2012). Dangers arise from a plethora of security problems including unintentional misuse of data, nefarious intentional abuse for the sake of financial gain, inattention, and from lack of understanding regarding the importance of data or proper security behavior (Dlamini et al., 2009; Leach, 2003). Personal data compromise creates risk of significant problems for individuals, among them "impersonation, fraud and identify theft" (Alessandro Acquisti, Friedman, & Telang, 2006)Personal data presents another danger to security, with evidence to support that personal data has, at times, been used to compromise high profile accounts (Reeder & Schechter, 2011).

Fiscal losses due to security breaches were estimated at $20 billion per year in 2000 in the United States and the cost of security breaches appears to be increasing. One study describes the global fiscal costs of cybercrime as $114 billion annually (Alazab, Abawajy, Hobbs, & Khraisat, 2013). Furthermore, proper security necessitates costly investment in infrastructure in addition to the cost of recovery and the development and growth of a

knowledge economy necessitates information security (Caldwell, 2014; Muhammad, Garba Ali, & Iliya, 2015; Tam et al., 2010). Security breaches may also impact other financial measures such as overall company value and stock value (K. Campbell, Gordon, Loeb, & Zhou, 2003). Individuals, as well as corporations, are deeply affected by security breaches. Real and potential harms include fiscal costs and psychological harm, among others (Caldwell, 2014; Calo, 2011; Toe, 2013).

Clearly, security is vital to continued fiscal wellbeing to nations, organizations and individuals. This security is compromised in a variety of ways, but for the purpose of this study, may be considered threatened by the disastrous combination of the human need for memorability in passwords, the design of many secondary authentication mechanisms and the ready availability of personal data used in authentication.

1.6 Research Questions

The following questions seek to explore the problem of available personal data frequently used in authentication and secondary authentication and provide a meaningful and usable analysis of data availability, the ease of access, and accuracy of personal data commonly used in authentication.

1. What personal data used frequently in password creation and secondary authentication are most likely to be identified correctly by a stranger using web resources?

2. How difficult is it for participants to find and correctly identify the personal data of a stranger that is used commonly in authentication using information available to them online?

3. Where is the personal data frequently used in password creation and secondary authentication most likely to be found by information seekers?

Chapter 2: Review of the Literature

2.1 Personal Data Protection and Definitions

Personal data are widely available online through a variety of sources, both legitimate and illegitimate. Regardless of the origins of data, it is widely believed that the extensive availability of personal data online and interaction around that data, are creating fundamental changes in social and technical environments. These changes are influencing refined definitions of privacy and security that reinforce the complexity of these concepts in the current context. Privacy and security are now discussed conceptually from the information systems literature in the context of social networks, online communities, and current legislation.

2.1.1 Privacy

Definitions of privacy vary widely, but may be broadly described as the ability to control and direct data access and use, particularly of personal data, to specific individuals or business functions. Foundational American legal understandings of privacy include the idea of the "right to be left alone" postulated in 1880 by Judge Tomas Cooley, Warren and Brandeis's foundation work in the Harvard Law Review in 1890, and subsequent legal definition in the United States Constitution in the Fourth Amendment and its ensuing interpretation (Palmer, 2011; Powell, 2011). Despite much attention, both in legal and information systems fields, the definition of privacy remains vague with a plethora of opinions regarding is actual implications, ethical and legal (Borchert, Pinguelo, & Thaw, 2014; Hartzog & Stutzman, 2013; Palmer, 2011)

Particularly speaking to information systems, Belanger and Crosser (Belanger & Crossier) describe privacy in terms of control, the ability of the user to control information about themselves including how data are disseminated. Discussions centered on control

continue in more recent discussions (Walrave, Utz, Schouten, & Heirman, 2016 & Heirman 2016). Hartzog and Stutzman (A. Acquisti, Adjerid, & Brandimarte) suggest that the term privacy is vague and misleading, since many activities conducted via information systems, such as posting information to social media are inherently public to some degree. However, it is vital to realize that users are often unaware of who may have access to information that they post, even if they take measure to protect their privacy (Madejski et al., 2012). Privacy policies change and so unintentional, from a user perspective, privacy breaches may occur (Gerlach, Widjaja, & Buxmann, 2015). Furthermore, the argument that people post data themselves does not absolve companies or organizations from responsibility to the consumer regarding privacy when policies change or systems are breached. Gerlach et al. (2015) argues that the right to privacy is not only a personal right but also a public good.

Instead of privacy, Hartzog and Stutzman (2013) suggest the term "obscurity", meaning "information is relatively difficult to find or understand" and a clearer and more attainable goal for designers. This definition seems limiting and is, in itself, vague and complicated in its application to information systems because "information that is difficult to find or understand" is the antithesis of many of the goals of user interface design and even of the broader concepts and goals of information systems in general. Another understanding of privacy from the perspective of positive user intent has also been suggested by Belanger and Crossier (2011). Privacy should reveal what the user wishes to reveal, only to whom they wish to reveal it, only for the user's desired duration it while simultaneously protecting against human error in either intentionally or unintentionally over-revealing information that may be harmful either to an organization or an individual, depending on context (Belanger & Crossier, 2011). Schneier (2010) describes privacy

abuses and the intentional manipulation of privacy settings and the design of interfaces to draw attention away from privacy. Control over secondary use, error, and improper access are also factors involved in privacy (Smith, Milberg, & Burke, 1996).

Clearly, the term privacy is undergoing a metamorphosis as attempts are made to adjust to the globally connected, data driven realities of the current information systems environment. Privacy becomes an important concept in the context of this study. A lack of privacy, through data compromise whether intentional through sharing sensitive data, or unintentional as in the case of misguided privacy settings or corporate data threat, may compromise the security of systems. The protection of privacy necessitates the security of systems which are most frequently guarded by passwords and secondary authentication that is, all too often, based upon personal data. Thus, a loss of privacy is both the potential cause and effect of security compromise.

2.1.2 Security

While not interchangeable, the concepts of privacy and security are certainly related conceptually in their mutual concerns. Security concerns itself with a different nuance of data use and includes a variety of descriptions. Security, like privacy, seeks to insure that only authorized persons are able to access and understand information (Dlamini et al., 2009). Security includes guarding physical aspects of information such as hardware and written records as well as insuring that only authorized individuals have access to information (Dlamini et al., 2009). Like privacy, this is a highly-nuanced concept, subject to change over time.

Ng, Kankanhalli and Xu discuss information confidentiality, essentially highlighting the use of data beyond the intended use, emphasizing the loss of control of

data (2009). This loss of security may occur within a particular organization by members of that organization who are authorized to access or use certain data, but only in particular contexts. A common example might be maintaining security of healthcare records only for particular purposes (Hall & McGraw, 2014). Questions regarding ownership and use of data will require broad definition and safeguarding in the form of legislation (Schneier, 2010). The substantial questions of legitimate data ownership and use that will inevitably become a matter of both national and international law are beyond the scope of this work, but remain an important part of the security discussion. The actual determination of who should retain access to information through the development of and adherence to security policies is also a significant concern to security. Human error and policy failure are frequently the causes of security failure in healthcare settings compared to the more researched technological failures (Rada, 2008; Sasse et al., 2001).

For purposes of this work, security is considered in the context of authentication, which is retaining control over the access to data by insuring that authorized individuals are correctly identified and provided with all the information which is within their prevue and none of the information that is not. Determining specifically which data individuals should access is a highly contextual question and must be addressed contextually in the literature. However, determining authentication, or positively and correctly identifying individuals is a question which spans contexts and challenges every area of information systems.

2.2 Authentication

Numerous attempts seek to insure that only authorized individuals are able to obtain access to systems and despite abundant recommendations, none of these alternatives

replace the much maligned passwords in common use (Bonneau et al., 2012). Other protocols that use two part authentication or other types of authentication often include a password-like component (Brown et al., 2004). All authentication protocols, from the most common password based system to the most complex multi-factor authentication have in common the motivation of maintaining privacy and security by ensuring access only to authorized persons.

Authentication schemes can be broadly divided into the following categories: possession, being, knowledge (Brown et al., 2004) . Possession describes authentication through the custody of an agreed upon object, such as a key card. The individual in possession of the token is provided access and, in absence of other authentication techniques, is assumed to be a valid user. “Being” or “existence” describes authentication on the basis of one’s characteristics. This category includes a variety of authentication methods such as fingerprints and retinal scans. Knowledge schemes rely upon a shared knowledge between the authenticator and the user and include passwords as well as password alternatives such as graphical passwords. There are certainly other approaches to classification of authentication within the security community, such as the one suggested by Park, Boyd, and Dawson (2000), however, Brown’s summary provides a meaningful framework for discussion in the current context (2004). In their extensive work, Bonneau et al. (2012) provide an detailed analysis of web authentication categories as well as specific examples of those categories as authentication schemes comparing their relative strength to password authentication in three areas: usability, deployability and security. Both these works provide a broad framework for authentication with specific development from other current work in authentication discussed as well.

Authentication protocols are presented briefly, as broad categories, and described in terms of general limitations in terms of the primary focus of this work: combined usability and security. At times, specific protocols are referenced to help describe the category. While appreciable differences certainly exist among the categories and the schemes, they benefit from organization into the four broad categories described above so that general characteristics may be discussed as they pertain to usability and security. It should be noted that not all the specific schemes clearly fit into a single category, for example, two categories may be leveraged by a single scheme. In this case, a primary category is selected with a notation of reliance on another category.

As this work seeks to augment current password research, a broad overview of the current authentication research will serve to provide a reference to current research and provide justification for continuing to improve passwords. To this end, current opportunities and limitations in authentication protocols are highlighted. Finally, the current state of passwords as the primary authentication protocol in current usage is discussed.

2.2.1 Knowledge Authentication

Shared knowledge is a key element to many authentication schemes, including passwords. Passwords clearly fall within this category and are probably the most representative type of knowledge authentication, passwords are discussed in depth along with other forms of knowledge based authentication. Also discussed are protocols that seek to improve either upon the usability or security of traditional password authentication are discussed. In knowledge authentication, the user, and supposedly only the user, shares a key piece of data in common with the authenticating system. Other technologies leverage

knowledge based security in a variety of ways often attempting to improve on the usability of passwords alone. Four approaches to knowledge authentication are now discussed.

2.2.1.1 Password Authentication

Passwords are probably the most well-known type of knowledge authentication or even authentication in general. A supposedly secret combination of characters is supplied by the user and verified by the system to gain access. Password protocols date to the earliest of authentication attempts and are still the most widely used authentication scheme (Bonneau, Herley, Van Oorschot, & Stajano, 2015; Herley & Van Oorschot, 2012). Passwords are frequently accompanied by expectations of users, such as the creation of passwords that adhere to certain standards (Shay et al., 2010). Research reveals various levels of adherence to and knowledge of these standards by users (Duggan et al., 2012; Tam et al., 2010). However, passwords have always, and continue, to challenge users in terms of usability. Users struggle to create secure passwords, which must be random, in light of the human reliance on pattern for memory (Baddeley, 1994; Brown et al., 2004).

2.2.1.2 Federated sign on

Password managers for web authentication allow the password manager to save the password for the user, accessible through a password used by the password manager (Bonneau et al., 2012). Federated sign-on, or single sign on schemes also rely on a single point of authentication, improving on the need for users to remember a multitude of passwords thus improving usability (San-Tsai et al., 2013). Examples of a federated sign on schemes include OpenID, Microsoft Passport and Facebook Connect.

Several popular single sign on protocols have also been found to be subject to significant security flaws (Armando et al., 2013). Federated approaches are subject to a

"weakest link" difficulty as a single compromised account compromises access to all the attached accounts. Federated approaches save account information should a user fail to log off the federated sign on, then all their accounts are likewise compromised. Should the password manager itself be compromised and account data stolen, a vast array of passwords for various systems and a multitude of users would be compromised. Mangers and federated sign on solutions improve usability, but on the security/convenience continuum sacrifice security to a degree which will be untenable to some.

Single sign on solutions also are ineffective in certain work environments. Heckle and Lutters (2011) found that in the collaborative clinical environment the single sign on approach was subject to increased security vulnerabilities and therefore suffered from resistance to implementation as users recognized these vulnerabilities. It has also been argued that part of a password's role is to engender trust on the part of the user and that federated sign on approaches do not enjoy the same level of user acceptance and trust as users (Biddle, Chiasson, & Van Orschot, 2012; San-Tsai et al., 2013). Thus, while federated sign on approaches improve usability in one respect, they also may not be applicable to some collaborative environments and they subject users to increased security vulnerabilities in several ways.

2.2.1.3 Graphical passwords

Graphical passwords rely on human memory of images, believed to supersede text based memory, and generally improve usability (Chiasson, Stobert, Forget, Biddle, & Van Oorschot, 2012; Furnell, 2007; Gao, Jia, Ye, & Ma, 2013). To authenticate using Persuasive Cued Clickpoints (PCCP), an example of the graphical category, users click on previously specified areas of five sequential images as a form of shared knowledge

(Bonneau et al., 2012). Similar protocols call for users to remember images, shapes, sequences and other types of visual patterns (Gao et al., 2013).

While graphical passwords show some promise in terms of human cognition and memory, users may also struggle with graphical passwords when required to remember multiple graphical passwords (Biddle et al., 2012). Additionally, Chiasson et al. (2012) discuss the "guessability" of PCCP the vulnerabilities of "hot spots" rendering graphical types of passwords susceptible to guessing attacks. Other graphical passwords are subject to both user error rate and security weaknesses that are often not fully explored in studies (Biddle et al., 2012).

2.2.1.4 Knowledge Authentication Summary

Clearly, some knowledge based authentication schemes improve on passwords in some areas of usability. A password manager improves on the usability of passwords in that only one password must be remembered by the user and in real use, the multitude of passwords which must be remembered by the user is certainly a hindrance to usability (Grawemeyer & Johnson, 2011). However, graphical and federated approaches to knowledge based authentication simultaneously introduce security risks and usability compromises.

2.2.2 Possession Authentication

Authentication based on possession of an object forms another subcategory of authentication schemes. Protocols are categorized as possession if they require the user to maintain a physical object in their possession, particularly if it is not an object they would be likely to carry commonly. Many possession authentication protocols rely on a combination of a possession and knowledge, such as an ATM card and pin. These types of

protocols fall within both categories, but are discussed with possession authentication. It should again be emphasized that this discussion represents a broad generalization of authentication mechanisms with specific examples for illustrative purposes designed to stimulate discussion around the concepts of usability and security.

2.2.2.1 Paper tokens

Examples of possession based authentication include some "proxy" type schemes, such as URRSA, where users authenticate through "'a man in the middle' between the users machine and the server" (Bonneau et al., 2012). URRSA relies upon hard copies of single use codes rendering it a possession based authentication. The physical use of the token along with the short-term use of the codes provides a security improvement on passwords as the user themselves do not generally, nor are they expected to, memorize codes. Randomness is increased as a result of eliminating the need to memorize and complexity can be increased as well, improving security. However, loss of the paper token could be catastrophic to security. OTOV bypasses this problem by requiring both a password and a code, however, this increases usability problems (Bonneau et al., 2012).

2.2.2.2 Software tokens

Adams and Dimitrinu (2008) propose a two-factor authentication scheme reliant on software tokens and cell phones that eliminates several problems with possession based authentication, primarily the need to carry a separate physical token. However, this also introduces possible problems such failure in the case of device power failure (Adams & Dimitrinu, 2008). Also, because the solution provided is a software based token, it lacks the increased security provided by a physical token.

2.2.2.3 Hardware Tokens

Hardware tokens, like their paper counterparts, may be used to attempt to increase security by requiring the presence of a physical object. These schemes necessitate carrying equipment that will provide all, or part, of the authentication. For instance, RSA SecurID relies upon a module in possession of a secret "seed" which generates a six digit code every minute. The current code is displayed to the user who types their user name and a "passcode", a combination of a user selected four-digit pin and the randomly generated code, to access a system or facility. While possessing an impressive track record, in March 2011, attackers were able to access the back end system and predict codes generated by the tokens (Bonneau et al., 2012). While perhaps more secure than passwords alone, hardware tokens are clearly not completely impervious to attack while simultaneously sacrificing usability by virtue of requiring the user to maintain possession of an object, and frequently, to retain a knowledge component as well.

Another example of hardware authentication is the use of the mobile phone for authentication. Usability wise, this may be an improvement on paper and other hardware token schemes, since it only requires the user to maintain possession of an object which they are already likely carrying. In Phoolproof, one instance of mobile phone based tokens, the user must use a password as well as their mobile phone to use the system (Bonneau et al., 2012). This scheme certainly has merit from a security standpoint – in addition to obtaining a password, a hacker would also have to physically obtain the user's phone, a task which would be more difficult to execute on a large scale compared to obtaining passwords or keys from a back-end system. However, it is still reliant on passwords and introduces a new usability problem – users who misplace their phones cannot easily regain

access to their accounts (Bonneau et al., 2012). This limitation is true of all authentication schemes that rely upon the physical presence of hardware. Maintaining possession of a device which tends to be commonly carried, such as a mobile phone, may be superior from a usability perspective compared to schemes which necessitate carrying a separate device.

2.2.2.4 Possession Authentication Summary

The result of a combination of physical token and knowledge authentication improve security compared to a password alone, but less useable than either physical token or knowledge authentication alone, requiring the user to possess both the token, and remember the password. Unfortunately, the existence of the paper token also poses a risk in itself a physical token could be stolen or copied (O'Gorman, 2003). Additionally, since the password portion of combined protocols are subject to the security weaknesses of passwords in general, while more secure than only password authentication, paper tokens pay a high price in usability for their relative security gains.

In general, possession based authentication benefit in regards to security but also suffer from decreased usability. In addition to suffering from the same, or similar, usability problems as passwords, this approach introduces additional usability problems since users are required to maintain possession of a physical device or printed codes. Furthermore, the existence of physical tokens or printed codes introduces a new security risk, the possibility for new methods of attack, such as reproducing a paper token, stealing a physical token, such as a phone, or reproducing the algorithmic method used to increase the randomness of the authentication scheme, as seen with RSA SecurID.

2.2.3 Existence Authentication

The concept of existence authentication is authentication based on who one is, rather than what one knows (knowledge authentication) or what one has (possession authentication). (Brown et al., 2004) suggest that security systems designed around "what you have" . . . "and you who you are" may someday supersede traditional password based authentication schemes as technology advances allow these types of authentication to be used more universally. Existence authenticators rely upon physiological characteristics such as fingerprints.

2.2.3.1 Biometric Authentication

Biometric authentication relies upon physical characteristics to provide authentication and may be the most usable form of authentication with relatively little effort from users. (Bardram, 2005). Jain, Hong, and Pankanti (2000) provide a summary of eight biometrics in use including fingerprint, iris and retinal pattern scanning, signature, voice, face or hand geometric, voice print and facial thermogram. While new biometrics continue to emerge, as a "static" metric, biometrics are subject to several usability and security concerns worth noting.

While some physical effort may be required in the use of some types of biometrics, it has been postulated that use of biometrics would require a lower level of cognitive engagement and memory retention compared to passwords. However, the actual usability and cognitive demand of biometric authentication is understudied (Bardram, 2005). More resent research is pointing out that cognitive dissonance and task disruption are introduced with biometric authentication and that high levels of user engagement are still required (Piccolotto & Patricio, 2014; Trewin et al., 2012). Furthermore, the assumption of usability

has been challenged, particularly in the context of highly mobile environments (Bardram, 2005). Thus, while biometrics were heralded as the effortless method of authentication, the practical limitations of usability are becoming clearer.

There are times when it may be difficult or impossible for an individual to use a particular form of biometric authentication and, as a static authenticator, it is often impossible to recover from a loss. Drahansky, Brezinova, Hejtmankova, and Orsag (2010) point out that various skin diseases preclude the use of fingerprints for biometric authentication. The same is reasonably believed to apply to other forms of authentication which may be either rendered inaccessible based on individual physical characteristic or altered by physical conditions or injury. In addition to an inability to compensate for an inability to use a given biometric protocol, a loss of an ability to authenticate cannot be recovered (O'Gorman, 2003). Should, for example, a user unfortunately loose a finger or an eye, for fingerprint or retinal scanning respectively, recovery of the identity would be very difficult.

Second, biometric authentication, while attacked differently than passwords and other authentication schemes, is still subject to attack. O'Gorman (2003) argues that because biometric data are not a secret, it is exposed to different vulnerabilities than "secrets" based authentication Fingerprints may be lifted from glass and voices recorded or an individual's face may be photographed in a public setting (Bonneau et al., 2012; O'Gorman, 2003). While it would be drastically more difficult to acquire an physically replay a retina scan, it would not be impossible, in some systems, to capture and replay biometric data (Faundez-Zanuy, 2004). Biometric data, in its digital form, if captured and sent through an authentication system, can be very difficult to distinguish from an authentic

version. Finally, if biometric data are to be used on a mass scale for authentication it must also be stored. Storage of biometric data introduces the possibility that this data may also be stolen, compromising individual identity and creating the possibility of great difficulty in accurately re-identifying individuals correctly. Furthermore, once compromised, a new biometric identifier cannot be created.

2.2.3.2 Biometric Authentication Summary

Biometrics, while promising in many ways, also present great difficulty and are particularly concerning in regard to identity recovery if a system were somehow compromised. The potential for loss with biometric systems seems greater. While most other types of authentication can be altered when compromised and thus secured, protocols based on biometrics must be considered entirely and permanently compromised for a given individual on a given biometric factor once that factor is compromised. Biometric protocols must therefore thoroughly address these concerns before confident and consistent deployment may be enacted.

2.2.4 Social Authentication

Social authentication recognizes the unique way an individual interacts with society and relies upon the self-regulation of society and the existence of an individual within that society to verify their identity. Social authentication represents a unique and relatively new form of authentication on the basis of *being* within society. Essentially, the user is authenticated through either their own social knowledge or other individual's knowledge of them. While it is sometimes considered "knowledge" type authentication, social authentication is not entirely appropriately limited to a knowledge framework. Particularly, the unique characteristics of knowledge measured by social authentication differ greatly

from more traditional knowledge authentication schemes. Password and similar knowledge authentication schemes require the use of a shared secret and while social authentication also relies on shared information, the social verification process is not as reliant on a memorized set of shared facts, but on the knowledge on obtains by being who one is – that is, one's existence in society provides intrinsic knowledge that is needed to complete the authentication process (Yardi, Feamster, & Bruckman, 2008). The user is not expected to memorize new information, but, ideally, should possess the information as a result of their existence. Therefore, they type of knowledge required by social authentication is fundamentally different than that in other knowledge based frameworks. Some social authentication schemes rely not on who you know, but on who knows you, or trustee methods, further removing this type of authentication from knowledge based authentication methods (Gong & Wang, 2014). In this instance, social authentication is clearly not a knowledge based authentication from the authenticated user perspective. Two approaches under consideration for socially based authentication, social knowledge and trustee authentication are now discussed.

2.2.4.1 Social Knowledge Authentication

One is to ask the user to provide information about individuals they "should" know. This approach to social based authentication is not novel and is used by significant social networks such as Facebook (Rhee, Kim, & Ryu, 2009; Yardi et al., 2008). Social networks seem ideally suited to evaluate authentication on the basis of social knowledge. One study suggests that through the tagging process inherently built into Facebook and other social network sites, users can be passively authenticated through there correct identification of individuals in photographs (Yardi et al., 2008).

Social knowledge authentication is subject to several types of attacks. Face recognition authentication protocols, such as Facebook's, may be circumvented through the use of facial recognition software and public data (Polakis et al., 2012). Social engineering to gain access to "friend" privileges, including viewing photos of friends may also supply an attacker with additional knowledge, enabling them to outwit the authentication scheme. (Polakis et al., 2014; Polakis et al., 2012).

2.2.4.2 Trustee Social Authentication

Individuals may also be authenticated by others who know them, also leveraging human relationships and knowledge into a social form of authentication (Brainard, Juels, Rivest, Szydlo, & Yung, 2006; Gong & Wang, 2014). According to one discussion, Facebook's implementation of social authentications asks users to name people in three different photographs (Rhee et al., 2009).

Stuart Schechter, Egelman, and Reeder (2009) discovered a problem with trustee based authentication which is remarkably similar to the difficulty of forgotten passwords, that is, users forgot the identities of trustees. Trustees are also susceptible to social engineering attacks (Stuart Schechter, Egelman, et al., 2009). Problematic also is the volatile nature of relationships and the contextual nature of "friends" on Social Networking Sites (SNS) (Yardi et al., 2008).

2.2.4.3 Social Authentication Summary

Social authentication leverages some promising aspects of human relationship contexts and reflects, to some degree, natural human processes for authentication (Brainard et al., 2006). While this is a significant achievement, social authentication creates several disturbing opportunities to compromise authentication processes. Social authentication

requires a great deal more study to be commonly implementable and is significantly shares several flaws with passwords.

2.2.5 Authentication Summary

While many forms of authentication are currently under research, none of them has yet replaced the password in common practice. All authentication protocols suffer from unique challenges including security failures and usability compromises. Several protocols also share challenges in common with passwords as already discussed. It is now clear that there is no easy or quick solution to authentication and that as much as passwords have been maligned; they appear likely to remain in use for some time. Because of this reality, it is important to understand the implications of continued password usage and seek to mitigate the harmful consequences of security breaches. Of particular interest to this study is the reliance of authentication protocols on personal data.

Table 1 Summary of Authentication Protocols

Knowledge	*Possession*	*Existence*	*Social*
Passwords	some Proxy schemes	Biometrics	Social knowledge
Password managers	Paper tokens		Trustee
Federated, single sign on	Hardware tokens		
Graphical			

2.3 Personal Data in Context

Personal data becomes a significant threat to privacy and security when the personal data used in authentication is compared to the personal data available online. This availability, described in current literature is discussed in the context of the actors involved in data availability. Furthermore, the limits of current knowledge, as revealed in the literature, are discussed.

2.3.1 Authentication and Personal Data

Several authenticators share a particular common weakness; the reliance of the authentication mechanism on personal data. Many primary and secondary authentication schemes are deeply reliant on personal knowledge and it therefore becomes necessary to evaluate the vulnerability of authentication schemes to personal data knowledge publicly available on the web. Of particularly interest is an understanding of what relationship, if any, exists in connection between the information publicly available online and the data used in the authentication process.

Several of the authentication mechanisms already discussed are vulnerable to personal data flaws. Possession approaches mostly avoid personal knowledge data attacks, although personal knowledge of individual habits such as were an individual might locate personal effects, such as a wallet, can lead to theft of token. However, this type of personal knowledge vulnerability is considered less likely because of the effort involved (Bonneau et al., 2012). Therefore, risk is often considered relatively low for most types of accounts, however, also because of the effort and expense involved, most commonly used systems do not consider possession authentication a worthwhile precaution. However, some types of biometrics, many knowledge and social approaches to authentication are vulnerable to personal data attacks because of the types of data used to authenticate. Biometrics authenticators at risk include facial recognition software, which may be deceived in some instances by a photograph and because they are reliant on data not inherently considered private (Qinghan, 2005). Knowledge authenticators, such as passwords, often rely upon user created passwords to improve memorability. However, users frequently rely upon personal data to create passwords, thus creating a vulnerability to others who possess

personal data regarding the user. This vulnerability has already been demonstrated regarding secondary authentication approaches and will be further explored regarding primary authentication in this work (Schechter, Brush, et al., 2009; Toomim, Zhang, Fogarty, & Landay, 2008) Social approaches are also vulnerable to personal knowledge as social knowledge may often be acquired through tagged photos, relationships revealed through public social media pages, facial recognition software and may also compromise other individuals associated in a trustee or social authentication protocol (Gong & Wang, 2014; Polakis et al., 2014).

It is readily apparent that many authentication protocols are vulnerable to personal data compromises. Because it is not possible to evaluate every form of data used in authentication for public availability in one study, attention is now turned to passwords as the most common protocols currently in use (Bardram, 2005). Secondary authentication, also a common approach to recovery of passwords, will also be considered in the study.

2.3.2 Passwords and Personal Data

Password generation often provides a point of system vulnerability (Adams & Sasse, 1999). Studies in human behavior in password creation and management reveal users seldom set out to create a security conscious password and are instead more concerned with memorability (Brown et al., 2004). Users take this approach despite training and awareness, although it appears dependent upon the strength of training and support as well (Weirich & Sasse, 2001). Policy initiatives may have some impact on password selection, but according to some studies, policy initiatives do not significantly reduce the inclusion of personal information, such as names and birthdates, in password

creation (J. Campbell et al., 2011). Users may simply be assigned passwords, but this complicates the problem of password memorability.

Many studies agree that purposefully obscure passwords are relatively rare (Duggan et al., 2012). Although not all studies distinguish between the relative volume of personal data used in password creation, Brown et al.'s summary suggests that names, such as the names of self, pets and relatives, comprise a sizeable portion of passwords created ranging from around 30-78% in various studies (Brown et al., 2004). According to Brown, et al. other personal data such as dates, phone numbers, ID numbers and addresses also make up sizeable portions of passwords created by users (Brown et al., 2004). Names, significant words, addresses and similar personal data also enjoy a long history of use in password creation and have been confirmed repeatedly in research as significant components of password creation (Dhamija & Perrig, 2000; Morris, Thompson, & Gaines, 1979). The use of personal information in password creation is so well recognized that users are cautioned against revealing the origins of a password, such as its reliance upon a family name (B. D. Medlin, 2013). Additionally, users tend to reuse the same set of significant words, for example, if the name of a family pet is used for one password, it is likely that a name of a previous pet is also used as a different password (Taiabul Haque, Wright, & Scielzo, 2014). This tendency towards reuse, or modified use may also compromise account integrity.

2.3.3 Secondary Authentication by Question and Answer and Personal Data

Passwords, as already discussed, are not well suited to human use from a cognitive perspective. This weakness creates another system vulnerability which is also exploitable with personal data available online, potentially even more easily than passwords using

personal data (Brown et al., 2004). Personal data also appears prominently in secondary authentication by knowledge question. Studies of secondary questions actually in use reveal a strong reliance on personal knowledge, which may be unwittingly compromised when that knowledge is publically accessible (Bonneau, Just, & Matthews, 2010; Reeder & Schechter, 2011).

The strength of security questions as well as the ability of friends and acquaintances to correctly guess answers to security questions is explored in several studies (Rabkin, 2008; Reeder & Schechter, 2011; S. Schechter, Brush, et al., 2009; Toomim et al., 2008). The ability of strangers to guess or discover the answer to questions is also anecdotally reported in several high profile incidents (Bonneau et al., 2010). The extent of danger to strangers from similar guessing attacks, particularly when not regarding famous and public individuals, is not well known.

2.3.4 Summary of Personal Data Used in Primary and Secondary Authentication

Specific examples of personal data as well as their primary use in primary and secondary authentication are provided to allow for specific comparison with personal data that is available publicly online, although no known studies currently and directly correlate the two. *Table 2.2* provides summary of several types of personal data used in password creation and secondary authentication questions. While not exhaustive, the summary is generally representative of current knowledge of personal data used in password and secondary authentication by question in the reviewed literature.

Table 2 Examples of personal data used in authentication

Personal Data Type	*Authentication Uses*	*References*
Name example: mother's maiden name, a nickname, children's names, pet's names, middle name.	Primary and Secondary	(Brown et al., 2004; Dhamija & Perrig, 2000; Furnell & Zekri, 2006; B. Dawn Medlin & Cazier, 2005; Rabkin, 2008; S. Schechter, Brush, et al., 2009)
Number example: phone numbers, dates, id numbers	Primary and Secondary	(Brown et al., 2004; Dhamija & Perrig, 2000; S. Schechter, Brush, et al., 2009)
Place example: cities, numbers from addressess, street names, complete addresses	Primary and Secondary	(Brown et al., 2004; S. Schechter, Brush, et al., 2009)

2.4 Personal Data Availability

The use of personal data in authentication raises questions about the availability and discoverability of personal data facilitated through the web. The widespread availability of personal data and the accompanying dangers to privacy and security are well documented (Dlamini et al., 2009; Oravec, 2012; Pavlou, 2011; Schneier, 2010). Personal data are compromised in a variety of ways and by various actors (Benson, Saridakis, & Tennakoon, 2015). Actors in the context of the web may include individuals providing their personal information, what Benson et al. (2015) terms "first party actors" (Benson et al., 2015). Secondary and Ternary provider services, collect and redistribute data. Finally, data may be appropriated from any of the previous entities by actors with nefarious intent for the malicious use of data (Benson et al., 2015). Systems may also be regarded as actors in the context of the social web (Zeng & Lusch, 2013). Each of these actors may be involved, in various ways and to various capacities in compromising personal data (Benson et al., 2015).

2.4.1 Personal and Group Data Compromise

Personal data security compromise initially begins when personal data are shared by the user, either publicly, such as through a blog or privately with a second party such as a company for the purpose of obtaining goods or services (Benson et al., 2015). It is well established that user will often provide information for the sake of connection or another benefit (Vickery, 2015). Given the types of personal data used in authentication, as already discussed, the revelation of personal information may seem innocuous in the context of social interactions, but in reality represent a significant security danger (Reeder & Schechter, 2011).

It is often noted that individuals expressed concern about privacy is not readily reflected in their privacy behavior. In short, users, although expressing concern about privacy, this expression of concern is rarely observed to cause a user to stop using a system or not download an application to their cell phone (Sutanto, Palme, Chuan-Hoo, & Chee Wei, 2013). This trait has been observed in a variety of contexts including social and transactional contexts such as e-commerce (Kokolakis, 2015). In commercial contexts, users will compromise personal data for a better price on goods or other benefit and conversely, are only willing to pay for privacy in limited contexts (Berendt, Günther, & Spiekermann, 2005; Beresford, Kübler, & Preibusch, 2012; Norberg, Horne, & Horne, 2007).

The human regulation of information disclosure, as studied in communications theory, is a complex set of highly nuanced factors (Waters & Ackerman, 2011). Online sharing behaviors vary widely and may be influenced by a wide variety of factors (A. Acquisti & Gross, 2006; Beldad et al., 2011). Petronio (2002) postulates five privacy rules

governing personal revelation, including culture, gender, gender, motivation, context and risk-benefit ratio and which are observable in some contexts of social web mediated sharing as well (Waters & Ackerman, 2011).

Youn (2005) states that in some online contexts males are more willing to share information to receive benefits compared to females. In contrast, other studies were not able to find gender differences in some sharing behaviors, such having a Facebook account, but did note differences based on age and other social factors, such as status as a student (A. Acquisti & Gross, 2006). It is also postulated that females are motivated somewhat different motivations for sharing compared to males (Waters & Ackerman, 2011).

Many other factors influence sharing as well. Users may be influenced by a unique culture, which encourages sharing as well as by their own cultural contexts and beliefs (Cockcroft & Heales, 2005; Waters & Ackerman, 2011). Motivation factors, as already discussed, may vary between individuals based upon other factors but also are also generally believed to influence the willing to barter personal information for some type of benefit. Context is also critical to disclosure choices. For example, self-disclosure has been found to be greater using computer mediated communication compared to face-to-face communication (Schouten, Valkenburg, & Peter, 2009).

Unfortunately, the nuances of privacy that influence individual data disclosure are not always well reflected in the technology available for individuals to manage privacy (Ackerman, 2000). Social networking sites privacy settings have been found to suffer from usability problems which result in significant data sharing not intended by users (Liu, Gummadi, Krishnamurthy, & Mislove, 2011; Madejski et al., 2012). Relinquishing personal data incidentally may not be the individual's direct choice, for example, in the

event that a friend or acquaintance shares personal information. Unfortunately, once data becomes available to a second or third party, it is nearly impossible, in current contexts, for users to maintain any meaningful control over their personal data (Acquisti et al., 2013).

2.4.2 Secondary and Ternary Data Compromise

Copious personal information, for many Americans, is continuously collected and stored in many venues and may be compromised in a wide variety of situations (Il-Horn et al., 2007). Wide spread government and corporate use and publication of data also contribute to the amount of personal data available online. Research in e-commerce and related fields sought to influence users to greater willingness to share their personal information.

Personal information, acquired by second parties may be used for uses unintended by the original supplier (Smith et al., 1996). This may occur with or without permission from the originator and in a wide variety of consequences, including a breach in privacy. The result is that individuals completely lose control, or even the ability to control, their data, as the locations and uses of data are not always transparent (Ackerman, 2004; Custers, van der Hof, & Schermer, 2014).

In addition to secondary uses of data expose the user to potential harm through security compromises within those organizations. Like individuals' personal data, corporate and government data suffers from loss of control due to compromised accounts, at times, the very password and similar issues discussed. Depending on the type of corporate or government data held, compromised data may include seemingly innocuous data that could ultimately assist in a malicious attack on an account security.

Specific developments in the history of online social interactions, government applications, and commerce readily illustrate the problems of introduced by second and third parties. For example, the use of names online was altered dramatically by the implementation of social media, particularly the implementation of the requirement to use real names by several social media venues. This requirement altered the nature of online identity and also made individuals identifiable by name across platforms (Schau & Gilly, 2003; van Dijck, 2013). One study demonstrated links in social networks reveal private information that users do not wish to reveal, such as political affiliation (Lindamood, Heatherly, Kantarcioglu, & Thuraisingham, 2009). By extension, links may also reveal other personal information applicable to authentication. Although users may take steps to obscure specific data, such as the names of children, this information may also be revealed secondarily, for example, by other members of the social network (Schau & Gilly, 2003).

Name information such as mother's maiden name can also often be retrieved from public records (Griffith & Jakobsson, 2005). While recently, retrieving these public records would have required a visit to a physical location, digitization allows for relatively easy and often anonymous retrieval. One study found that pet's names were slightly harder to guess than human names, however, pet's names remain susceptible to statistical guessing attacks (Bonneau et al., 2010; Rabkin, 2008).

Many types of numbers are also commonly available online. Stutzman explored the information students posted voluntarily on social networks and found that numbers, such as birthdates, as well as addresses and name are commonly provided by users and even required by social networks (Stutzman, 2006). Location names, addresses, and names of organizations that supply location information can also frequently be derived from SNS

(Adamic & Adar, 2003). Business and healthcare records, either compromised or used beyond their original scope, also may potentially compromise all three areas if handled inappropriately (Hall & McGraw, 2014; Li, 2014; Sharma & Crossler, 2014). A summary of personal data exposure and the venues of exposure is provided in Table 3.

As a result of changes in information systems capabilities and practices, what were once considered offline identifiers, such as addresses, becomes online and digital identifiers (Preibusch, 2013). These digital identifiers make is possible to positively and accurately identify individuals. Combined with personal data provided by individuals and in the context of the personal data used in password creation, this data becomes dangerous to account security. Organizational policies and protection strategies have failed to keep pace with rapidly developing data storage and transmission capabilities rendering data available to the final actor in the discussion, malicious parties (Corbett, 2013; Hartzog & Stutzman, 2013; Powell, 2011).

2.4.3 Malicious Party Data Compromise

Information which is readily attainable, either publically available using government data, easily attainable from commercial sites or provided directly by individuals overlaps disturbingly with the personal data used in primary and secondary authentication (Reeder & Schechter, 2011). This can compromise information secured through the use of personal data in authentication protocols as attackers gain personal data used in password creation or secondary authentication. Furthermore, using the information, once obtained, is disturbingly effortless, particularly in the context of secondary authentication, where the common knowledge required for access is specified. This state

of affairs creates a danger to account security when personal data are acquired by malicious actors.

Several studies have sought to establish the level of shared knowledge between individuals and acquaintances, particularly as it pertains to secondary authentication as well as in the context of multiple access systems (S. Schechter, Brush, et al., 2009; Toomim et al., 2008). Guessing attacks have also been analyzed in the context of hackers using dictionaries or similar devices (Bonneau et al., 2010). However, the level of risk for individuals in the context of strangers is unestablished in terms of what information is intentionally discoverable in the context of a particular individual. This reality becomes significant in light of the personal approaches often used in attempts to compromise accounts. Social engineering, for example, that are involved in many security failures (Biddle et al., 2012; Kline, He, & Yaylacicegi, 2011). In the context of social media and personal data in secondary authentication, the social engineering could be as simple as creating a friend request or asking about pets on Facebook and using this information gaining access to an account.

Unfortunately, in addition to not assessing individual data vulnerability, it is also unclear how difficult it is to find data in the current online context. It is known from the literature that an over-abundance of data and information may obscure necessary information and make task completion more difficult in a variety of contexts (Amar & Stasko, 2004; Jones, 2004). Therefore, it is possible, that while information is known to be available, it may be too difficult to locate for practical purposes in the context of attempted security violations.

Studies in user password creation and analysis of secondary authentication questions as already described provides the ability to compare personal data used in authentication with personal data available online. Additionally, the locations of the data in question reveal that the problems of the availability of personal data are not simply a matter of privacy control for individual users, but a challenge to current data management in corporate and government sectors as well.

Table 3 Examples of personal data available online

Personal Data Type	*Data Locations*	*References*
Name example: mother's maiden name, a nickname, children's names, pet's names, middle name.	Social networking sites, public records, business records, health records	(Griffith & Jakobsson, 2005; Li, 2014; Lindamood et al., 2009; Sharma & Crossler, 2014; Stutzman, 2006)
Number example: phone numbers, dates, id numbers	Social networking sites, public records, business records, health records	(Hall & McGraw, 2014; Sharma & Crossler, 2014; Stutzman, 2006)
Place example: cities, numbers from addresses, street names, complete addresses	Social networking sites, public records, business records, health records	(Hall & McGraw, 2014; Sharma & Crossler, 2014; Stutzman, 2006)

2.5 Conclusion

There is a significant level of risk associated with the personal data used in authentication and the personal data which is publicly available. Abundant anecdotal evidence reveals the possibility of attacks on the basis of personal data which has become public (Bonneau et al., 2010; Reeder & Schechter, 2011). However, the level of personal exposure and danger to non-celebrity individuals may vary widely as information available differs between individuals. Furthermore, even when data exists online, it may or may not be discoverable to strangers.

The importance of maintaining privacy and security and the current understandings of privacy and security were provided. Many types of authentication are pervious to personal data attacks as already established, potentially compromising privacy and

security. Finally, the current literature on data availability is analyzed for similarities to personal data used in authentication and secondary authentication. This review of the literature reveals a gap in current knowledge addressed in the following chapter.

Chapter 3: Experimental Design

This multiphase study explores the availability of online personal data used in primary and secondary authentication protocols to a human information seeker. The phases of the survey each focus on data collection from two different groups of participants, the information source participants (source participants) and the information seeker participants (seeker participants). Key to the experimental design is a survey which provides the fundamental data located by the seeker participants in the experiment described in the following section. The survey is followed by a questionnaire provided to the source participants that establishes the accuracy of the personal data garnered from the survey. The data provided as a result of the complete experiment includes the following:

3.1 Independent Variables

The participants themselves are the independent variables in the study. The information source participants' personal data, found online and available to the source participants may be influenced by a wide variety of undetermined factors. The participants and their roles are discussed further in the during the *Procedure* and *Participant Recruitment* sections. The focus of this study will include assessing the availability, to strangers, of personal data used in authentication. An analysis of the location and difficult of retrieving the data are also provided as a result of the study.

3.1.1 Source Participants

Source participants are the independent variables contributing the dependent variables of personal data and potentially, although not definitely, influencing the location of ability.

3.1.2 Seeker Participants

Seeker participants influence the dependent variables of time, influence the location provided in the survey search, the time spent on the search and provide their personal perspective on the difficulty of retrieving information.

3.2 Dependent Variables

From the independent variable data gathered the following variables are provided for analysis.

3.2.1 Personal Data

The personal data used in authentication under current review is dependent upon the source participant. Source participants may influence the availability of personal data through their online sharing habits, but research demonstrates that online sharing is not the only source of personal data and thus, personal behaviors and privacy concerns may or may not influence the data that may be found by another individual (A. Acquisti & Gross, 2006; Griffith & Jakobsson, 2005). Personal data are collected by the seeker participant and compared to the accurate personal data as described by the source participant.

3.2.1.1 Personal data as provided by source participant

The answers to personal data questions, as supplied by the seeker participants, are confirmed for accuracy with the source participant following data collection by the seeker participants. This occurs through the use of a questionnaire containing a composite of all the answers from assigned seeker participants with the opportunity for source participant to confirm the accuracy of each unique answer provided by seeker participants.

3.2.1.2 Personal data as determined by seeker participant

The personal data of the source participants, as determined or guessed by the seeker participants, is recorded. The personal data are pursued by the seeker participants regarding an assigned source participant. The personal data under investigation is selected from IS literature regarding personal data frequently used in password creation and secondary authentication as described in Chapter 2 and are collected via an originally designed survey instrument modeled on available information from previous experiments described in the literature review. The seeker respondent has the opportunity to provide a single answer, multiple answers or "guesses" regarding the personal data, or no answer with an indication that an answer has been either attempted or not attempted.

3.2.2 Time Dedicated to Search

Each seeking participant provides an approximate estimation of the time required to obtain the answer to each question, or in the case of multiple guesses to a single question, the total time spent attempting to answer the personal data question. This information assists in determining the ease of correctly identifying the data. The time to complete a task is used in studies across a variety of IS topics, such as the usability of mobile applications (Hoehle & Venkatesh, 2015). Time is often considered as both a factor in usability and an influencing factor in the actual use of systems to locate information in a variety of specific fields ranging for physicians use specific information resources to consumers' behavior in electronic commerce applications (D. P. Connelly, Rich, Curley, & Kelly, 1990; Faiola, 2007). Here, time may be influenced by the difficulty of locating source participant information as well as by seeker participant behaviors and choices.

3.2.3 Location of Personal Data

Each seeker participant records the web location of the data as part of the survey. "Location" is requested for each guess made by the participant. The preferred data type is a Uniform Resource Identifier (URL). However, the seeker participant is asked to describe the location of the data if the URL is not available or impractical due to collection method (e.g. a written survey) where a URL may not be practical to record accurately. Descriptive reporting may include how data was derived or "guessed" by the user. Data locations may vary as seeker behavior is variable. Searching behavior and reliance on different types of media has been shown in research to be variable on a variety of factors (Kim, Sin, & Tsai, 2014; Qiong, 2013). Locations may also be influenced by source participant sharing behaviors.

3.2.4 Difficulty Rating by Seeker Participant

The seeker is asked to provide a Likert scale rating describing the difficulty or ease of locating information. The Likert scale is provided in the Personal Data Survey located in Appendix 1. Likert scale questions enjoy prominence in the social sciences and have been used in IS literature to evaluate topics such as user perceptions of privacy and security (Anderson & Agarwal, 2010; Gerlach et al., 2015; Lozano, García-Cueto, & Muñiz, 2008; Preibusch, 2013). Likert scale type questions are also used in the IS literature to help evaluate usability (Finstad, 2010). The question serves a similar purpose here: to evaluate the difficulty of using the various resources available to the participant to complete the task, in this case determining level of ease or difficulty associated with locating answers to personal data questions. This perception of difficulty is a factor in considering the difficulty

of locating personal data. A Likert scale is adapted and applied for use in this survey (Vagias, 2006; Wilson, 2013).

3.3. Procedure

To determine the difficulty of discovering personal data online, a multi-phase approach is applied. The study includes two primary data collecting phases following initial participant recruitment. In combination, these phases allow an evaluation of the research questions. In addition to multiple phases for the study, a pilot study is also conducted in order to more fully inform the methodology and determine the best procedure for the dissertation study. Descriptions of the pilot study, participant recruitment, and dissertation study follow.

3.3.1 Pilot Study

An initial pilot study with source participant and four seeker participants explored the implementation of the study. The procedure follows the same multiphase process as the full study, beginning with participant recruitment and informed consent. The experiment moves into the collection of the personal data using the survey instrument. Seeker participants identify specific data points regarding assigned seeker participants, the location of those data points, and the time dedicated to determining the data point. Finally, the seeker participants provide an evaluation of the difficulty of locating individual personal data points using a survey instrument. The experiment culminates in confirmation, by the source participants, of the accuracy of data gathered during the survey. The same procedure is followed for both the pilot and dissertation studies. A summary of the research questions, variables collected and addressed and instruments is provided in *Table 4.*

Table 4 Summary of Variables & Research Questions

Research Question	Independent Variable	Dependent Variable	Assessment Method
1. What personal data used frequently in password creation and secondary authentication are most likely to be identified correctly by a stranger using web resources?	Source Participant Seeker Participant	1 – Personal data 2 – Time 4 – Difficulty rating	Survey Questionnaire
2. How difficult is it for participants to find and correctly identify the personal data of a stranger that is used commonly in authentication using information available to them online?	Source Participant Seeker Participant	2 – Time 4 – Difficulty rating	Survey Questionnaire
3. Where is the personal data frequently used in password creation and secondary authentication most likely to be found by information seekers?	Source Participant Seeker Participant	3 – Location	Survey

The pilot study primarily informs two specific methodology choices. First, the pilot study provides an opportunity to evaluate the amount of data required for the seeker participant to correctly identify the source participant. Of particular interest is the necessity of including photograph of the source participant. One half of seeker participants are supplied with a photograph of the source participant as well as the source participant's name and city and state demographics. As described in the literature review, this combination of data are often sufficient to identify individuals on potential web resources such as blogs or Social Networking Sites. The remaining seeker participants are provided with only the source participant's name and city and state. If the source participant is

identified accurately by the seeker participants without the use of a photograph, through the use of name and city and state data alone, then only this data are provided to future participants to aid in the ease of study distribution and data collection.

The pilot study allows for the testing and refining of the original survey instrument, particularly with attention to the survey's distribution in a hard copy or as a web survey. While previous, similar studies made use of a web-survey design, it has been established in research that information processing may occur differently using physical artifacts, for example, reading comprehension and speed may be impacted by the use of screens compared to physical books, although some recent studies note that differences between paper and digital processing may be task and outcome specific (Dillon, 1992; Noyes & Garland, 2008). The pilot study provides half of the seeker participants with a physical copy of the survey and half the participants with a digital, web administered copy. The results are compared for completion rates and accuracy and the administration method yielding the most complete results are continued for the dissertation study, if an appreciable difference exits.

3.3.2 Dissertation study

The completion of the pilot study, providing initial insight into the research questions and particularly into the methodology for source participant identification and the method of administering the survey, affords improved decision making for the delivery of the survey and the data provided to seeker participants. Aside from these choices, the main study proceeds with participant recruitment, the seeker participant survey data collection, data accuracy confirmation with the source participants and terminate with an analysis of the resulting data.

3.3.2.1 Participant Recruitment

As already described, this study relies upon the use of two groups of participants. Seeker participants attempt to discover personal data regarding another individual, the source participant. Source participants allow others to attempt to identify their personal information online and then confirm the accuracy of the data provided by the seeker participant. Each group of participants is recruited independently. Descriptions of specific recruitment techniques and requisites that apply to each group of participants are addressed in the following sections *Source Participant Recruitment* and *Seeker Participant Recruitment.* Details of subject selection that apply to both groups are now discussed. Recruitment of participants for the dissertation study is conducted after completion of the pilot study to collect the full cohort of four source participants and thirty to forty seeker participants. The resulting full data from the seeker and source participant recruitment, which provides a total of one hundred twenty to one hundred sixty responses across four source participant's data reflects similar participant recruitment strategies and cohort sizes to other, similar studies such as Schechter, et al. study, which used one hundred thirty participants in pairs (2009). Informed consent will be obtained from both groups as participant recruitment occurs.

3.3.2.1.1 Source Participant Recruitment

Source participants will be recruited using referral based sampling beginning with the selection of several individual acquaintances of the researcher. Participants will be asked to recommend other individuals for participation. The research recruitment is designed to expand the study demographic well beyond university student recruitment and include a wider demographic than would be available through university student

recruitment alone. Prior to the first phase of the study, source participants are initially selected to minimize potential relationships between source participants and seeker participants and participants asked to refer other individuals who may be interested in participation in the study to expand the reach of the study. By recruiting outside the university, it is hoped that potential prior relationships between source and seeker participants will be minimized.

Four source participants are recruited to represent a variety of backgrounds, particularly with regard to age and gender. Two female and two male recruits from a variety of age groupings are recruited, reflecting recruitment patterns in similar studies (S. Schechter, Brush, et al., 2009). Pew's research study groupings for studies of social media use which include adults ages 18-29, 30-49, 50-64 and over 65 (Brenner & Smith, 2013). One adult will be recruited from each age group to represent a spectrum of online engagement and data exposure across various generations.

Source participants' engagement in a social network is not a requisite to participation in the study for either group of participants, as individual's information may be available online, even if they are not heavily participatory in online social networks or other methods of online communication, such as blogging. While Acquisti and Gross found that factors such as information privacy attitudes are linked to online behavior, online sharing is not the only source of information (2006). Prior study reveals that information is often available via sources where personal information is not directly revealed by the user such as public records (Griffith & Jakobsson, 2005). Because information may be available from sources additional to source participant's own online participation in sharing, online behaviors are not considered prior to selecting participants. As it will be

necessary to provide strangers with their biographical data, participant's willingness may be biased towards participants more willing to be scrutinized, and hence, potentially less concerned about security, which may influence outcomes, but until the locations of information provided for seeker participants are established, this is impossible to verify.

Informed consent is obtained at this time and sufficient biographical data collected to allow online identification by strangers, such as name, city and state, and photograph (Gross & Acquisti, 2005). The pilot study previously performed determines the amount of data provided collected from source participants and provided to seeker participants. None of the biographical data collected from the source participants includes data questions addressed in this study.

To mitigate both concerns of participants and actual risks of data exposure, participants are notified of the outcome of the study with regards to their particular personal information found online. Source participants are provided with a list of the data points seekers will attempt to ascertain. While this precaution may prompt a change in behaviors, it is considered necessary to protect source participants to provide them with the opportunity to change any passwords or security questions dependent on the data sought in the study. The researcher requests that the participant avoid changing their online behavior during the duration of the study and avoid changing any personal data available online at the time the study commences. While these precautions may again, influence the outcome of the study, the potential for harm from compromised accounts is a primary motivation for the study and protecting participants from such harm as a result of participating in the study must be avoided.

3.3.2.1.2 Seeker Participant Recruitment

Seeker participants are also recruited using referral based sampling methods, beginning with acquaintances of the researcher but purposefully avoiding individuals known to be acquainted with the source participants. Recruitment of undergraduate students is also used for seeker participants through the use of on campus posters and classroom recruitment. Informed consent is obtained from seeker participants. Seeker participants will be asked to refrain from any illegal methods of obtaining information. No known risks are involved for seeker participants beyond the fatigue involved with using computers for an extended period of time and, if using the written survey, the physical act of writing. Since time is not limited and the survey may be taken at their convenience, participants may rest as needed.

3.2.2 Personal Data Survey

To understand what strangers are able to discover about individuals, each seeker participant is provided a list of personal data points to discover about the source participants. The seeker participant was provided with the name and city and state of the source participants and asked to identify the source participant as a stranger, that is, someone with which they have not previously communicated directly either in person or online. The data provided the seeker participant included a photograph of the source participant, determined necessary in the pilot study. None of the biographical description includes any part of the personal data sought. For each point of personal data, the seeker participant provides the requested personal data and may supply multiple guesses, if desired (Schechter, Brush, et al., 2009). Additionally, the approximate time as reported by the seeker spent locating the data, the location of the data, and the perceived difficulty of

locating the data are obtained. The information was collected web based survey, as determined by the pilot study. Additionally, the web based survey allowed for timing the beginning and end of the web survey. The personal data points include commonly used questions for secondary authentication as well as the personal data most commonly used in passwords derived from currently available literature on the topics of password creation and secondary authentication. A description of the personal data used in authentication follows and included in the study follows.

3.2.2.1 Personal Data Criteria for Inclusion

The target personal data are derived from studies of both primary and secondary authentication models. Literature regarding primary authentication, as already discussed, reveals commonly used personal data points for primary authentication in the password creation process. Studies also reveal commonly used personal data questions from secondary authentication. Combining both primary and secondary authentication models reveals some overlap between the two groups as well as providing a set of personal data metrics to test for availability. Personal data requested will include the following. “Personal” data describes the data point of interest regarding the source participant; use describes the primary role of that data point in authentication, either primary (used for password creation), secondary (used as an account recovery question). As the majority of passwords created by users make use of names and names are also featured significantly in secondary authentication as described in the literature review, the identification of names is selected as the primary focus of this study.

Table 5 ***Personal Name Data Used in Authentication***

Personal Data	**Principal Use**	**Source(s)**
Mother's Maiden Name	Secondary	(Furnell & Zekri, 2006)
Nickname	Primary & Secondary	(Brown et al., 2004; Rabkin, 2008)
Child(rens) Name(s)	Primary	(Brown et al., 2004; B. Dawn Medlin & Cazier, 2005)
Pet(s) Names	Primary & Secondary	(B. Dawn Medlin & Cazier, 2005; Rabkin, 2008; S. Schechter, Brush, et al., 2009)
Middle Name	Secondary	(Rabkin, 2008)

3.2.2.2 Excluded Personal Data

While not every possible data point used in primary or secondary authentication is tested during this experiment, this procedure could be used to test other personal data points in the future. The data points selected represent a cross section of data points commonly used in primary and secondary authentication as described in the literature with an emphasis upon those points which capture the highest frequency and are preferably used in both primary and secondary authentication.

In this case, name data is emphasized as "names" questions including self, family and pet names, represent the sizeable majority of personal data used in password creation for primary authentication and a sizeable minority of secondary account authentication questions (Brown et al., 2004; Rabkin, 2008). Some points of personal data, such as an actual individual's name, could not be used practically, but are frequently used by individuals in password creation (Brown et al., 2004). The design of the study does not support the use of the source participants name as a data point as failure to provide a name would make it impossible for seeker participants to identify source participants.

Another commonly used personal data point in password creation is an ID number, such as a social security number, also described in the literature review. While any personal data point used in authentication is potentially damaging in that it could lead to compromised security of an account, many ID numbers like social security numbers, were deemed too intrusive and could potentially more easily lead to harm for the participants and are therefore excluded from the study.

3.2.2.3 Survey Instrument

Subsequent to identification, recruitment and informed consent of source and seeker participants, each seeker participant is asked to identify the information facts, as accurately as possible, about an assigned source participant. The participant is permitted to include multiple guesses, if desired. Schechter, Brush, et al. (2009) used this method to encourage more active searching on the part of participants in a similar study design. Participants are asked to include the location of the data in the survey. A URL or description of the location of the data are requested. The seeker participant is also asked to provide the approximate time required to locate the data as well as a Likert scale rating of the difficulty of providing the answer. Both a paper and web version of the survey were created to facilitate the collection of as much data as possible. A pilot study, already described, ascertained whether paper or web surveys obtain better results in terms of complete data and participation. In either instance, user may answer as many questions as they like and elect to leave some data points unanswered if they desire. They are however, asked to indicate whether they attempted a data point and the amount of time they dedicated to a search, if attempted. This measure distinguishes between unanswered and attempted questions and unsuccessfully attempted questions. Of course, in a paper survey, the

completion of all data cannot be controlled and participants were encouraged, but cannot be required, to supply all data requested[1].

It is also possible that some data may be acquired in searching for other data, for example, a participant may learn a current pet's name in the course of looking for a first pet's name and thus be able to answer the "current pet name" question immediately. Time for task information and data location will also be assessed during analysis to locate personal data where this may impact the accuracy of the total time for a personal data point question.

The Personal Data Survey, an original instrument, was administered as described. The survey is available in Appendix 1. As a result of survey completion by seeker participants, variables 1, 3, 4 and 5 are addressed. Question 1 in the survey provides the specific data guess or guesses by information seeking participants. Question 2 and 4, the

[1] One survey required similar guessing or searching of information used in secondary authentication questions by acquaintances of the survey participants and that instance a similar procedure was followed using a web-based survey instrument occurring on site, in separate rooms between pairs of participants to prevent collusion regarding answers(Schechter, Brush, et al., 2009). A lab controlled situation is not considered as necessary in this experiment as the participants are not acquaintances. Furthermore, a laboratory environment is restrictive to the time spent on the search (Schechter, Brush, et al., 2009). Here, the survey instrument is used in an uncontrolled environment to allow the searcher to dedicate as much, or as little, time as they wish to obtain the information. It is anticipated that in some instances, participants may lose interest in the survey and stop searching as diligently (Schechter, Brush, et al., 2009). This was mitigated with the use of incentives in the form of class extra credit as approved by the IRB and course instructors.

time required to locate each personal data point and the Likert scale rating are together used to evaluate the difficulty of locating data and the location of the data reveals points of vulnerability to information searching. As a result of the first survey, several of the research questions may be addressed as well. *Table 6*, provided in the description of the pilot study, and provides summary of the survey instruments, the variables addressed and the research questions.

3.2.3 Data Verification Questionnaire

During data verification, each source participant is asked to confirm whether or not the data provided by the seeker participants was correctly identified. A cumulative list is provided to each source participant of the answers provided by seeker participants. An example of a data verification questionnaire for a single data point is included in the Personal Data Verification Survey in Appendix 2. This questionnaire is individually crafted for each source participant on the basis of the data provided in the information seeking survey and cannot be produced in its entirety until data collection from the Personal Data Survey is complete. As a result of the data verification process, it will be possible to assess dependent variables for accuracy.

Source participants are presented with the questionnaire for each of the personal data points associated with the seeker participant survey. Source participants were previously encouraged to change any passwords or secondary authentication questions as needed and informed that this is the information other people were able to obtain regarding themselves. This strategy for confirmation will avoid revealing the answers to any personal data questions that the seeker participants were unable to verify, protecting any data that is not already available on the web.

3.2.4 Dissertation Study Summary

A summary of the research questions, the variables addressed and assessment methods appear in *Table* 6. A two-phase study is used to address the research questions. The pilot study allows for refinement of the survey instrument and the following dissertation study will address the research questions directly through the use of two groups of participants. The source participants will provide consent for select personal data to be located using web resources, if possible, by the seeker participants. The seeker participants will provide additional data about their search process. Following the conclusion of the search by the seeker participants, source participants will be asked to confirm the accuracy of personal data discovered. The data collected will provide a rich venue for the analysis of availability, accuracy, location and subjective assessment of the difficulty of locating personal data belonging to a stranger using web resources.

3.4 Analysis

This work seeks to understand the relationship between personal data used in primary and secondary authentication and the accuracy, location and difficulty of accurately locating personal data. The result is an understanding of the personal data points are the most vulnerable to discovery by strangers and the web locations that are vulnerable to compromise. The weakness of personal data in authentication is well established but likely to continue as discussed already. Therefore, an understanding of vulnerabilities is helpful for both recommendations for password creation and in the design of secondary authentication questions.

3.4.1 Availability and Accuracy of Data Located

In order to address the research questions, the availability of the data are assessed by the ability of seeker participants to identify personal data about the source participants accurately. This outcome is assessed using the personal data identified or guessed by seeker participants and affirmed by source participants. The correct and incorrect data collected are assessed respective to individual source participant to evaluate their individual vulnerability on the personal data points. Personal data points are also evaluated and reported across the sample to find the most vulnerable data points. These metrics will be provided in the form of descriptive statistics.

To assess data vulnerabilities, the statistical rate of a correct answer to a data point will be calculated. Any answer designated by the source participant as "correct" in the data verification survey will be considered a correct answer. Likewise, those designated incorrect by the source participant will be considered incorrect answers. After establishing correct and incorrect answers, statistical analysis is performed to yield the percentage of correct answers by source participants. These statistics may also be reported individually by source participant for comparison between source participants and correlation to demographic factors using ANOVA. Data accuracy is assessed by question, with a total reported for each question and by participant. A total percentage reported by each participant gauges the consistency between correct answers for the total group and reflects the vulnerabilities of individuals. This process will be helpful in assessing whether or not some data vulnerabilities are consistently more common than others.

3.4.2 Difficulty of Data Locating

The outcome of the data collection leads to a determination of the difficulty of accessing the information. The difficulty is evaluated comparatively between the different personal data points based upon the accuracy the determinations by seeker participants as well as the perception of difficulty and the comparison of time required for search. To analyze this data, time and correct data identified are correlated using ANOVA.

If information is correctly identified a similar percentage of the time, but one data point takes substantially more work, in terms of time, to identify, it might be more secure than easily identifiable personal information. To assess the difficulty of locating correct data, the correct data points will be compared with time and the Likert assessed difficulty scale. Time is often used as a metric in usability testing and here, is expected to assist in reflecting the difficulty of correctly locating data (Sauro & Lewis, 2010). The Likert difficulty scale will be observed for correlation to the time spent identifying the data using ANOVA (Ostertagova, Ostertag, & Kova´c, 2014).

3.4.3 Data Locations

The location data provided by the seeker participants and may be or may not be dependent on the online behaviors of source participants as already discussed. Location data may also be reflected in the search process of seeker participants. The location data are evaluated for consistency and differences between source participants to providing insight into areas of vulnerability in personal data on the web. The data used to address the research question regarding data locations are an accumulation and analysis of the locations provided by the seeker participant data. The data are expected to be reportable in the form

of a categorical analysis of the types of locations that data are available as well as a summary of specific web locations used for data discovery.

3.4.4 Analysis Summary and Research Questions

The analysis plans describe the expected approach to the data which will ultimately address the research questions. Each research question is reviewed with a summary of the data analysis which will seek to address the research question.

Table 6 ***Summary Research Questions & Analysis***

Research Question	**Instrument & Variables**	**Analysis Plan Summary**
1. What personal data used frequently in password creation and secondary authentication are most likely to be identified correctly by a stranger using web resources?	***Seeker Participant*** *Survey* 1 – Personal data 2 – Time 4 – Difficulty rating ***Source Participant*** *Questionnaire* 1 – Personal Data	Descriptive statistics analyzed by personal data question and source participant. ANOVA to evaluate correlation between time and accuracy. ANOVA to evaluate consistency between source participants
2. How difficult is it for participants to find and correctly identify the personal data of a stranger that is used commonly in authentication using information available to them online?	***Seeker Participant*** *Survey* 2 - Time 4 – Difficulty rating ***Source Participant***	Descriptive statistics analyzed by personal data question and source participant. ANOVA to evaluate correlation between time and accuracy.
3. Where is the personal data frequently used in password creation and secondary authentication most likely to be found by information seekers?	***Seeker Participant*** *Survey* 3 – Location ***Source Participant***	Descriptive statistics Post facto summary of common location types Analysis of variability across source participants

The multiphase study, as described, is anticipated to provide new insight into the research questions as well as to provide a foundation for future research which are discussed in the *Results* and *Future Research Directions.*

Chapter 4: Results

4.1 Overview and Analytical Process

4.1.1 Introduction

Results collected from both seeker and source participants with analysis as appropriate are provided in Chapter 4. The surveys were distributed and collected as outlined in Chapter 3. Following the collection of data, the results were reviewed by each source participant for verification of data accuracy. Analyses of the results are provided. Chapter 4 is divided into three sections: Introduction, detailed results reported in order of source participant and a summary of results considering the entire study.

The first section provides the introduction, details data cleansing and describes the analytical process. Data cleansing gives insight into the raw state of data, the necessary elimination of incomplete surveys and provides detail regarding the coding of data into accuracy groupings for comparison.

In section 4.2, results are reported as they pertain to each source participant. For ease, each of these result groups are reported in the same order. First, demographic information is provided for each source participant. This is followed by a summary and comparative analyses of the personal data regarding each source participant. Finally, each individual personal data point is analyzed individually with particular regard to understanding the location and difficulty of accurately identifying for that data point. For each data point, complete descriptive statistical analyses are provided along with frequency measures and ANOVA as appropriate. This process is repeated for each source participant.

In section 4.3, results are summarized across all four seeker participants to provide an aggregate representation of the accuracy, location and difficulty of discovering each

data point. This section provides analyses which highlight the similarities and differences in data accuracy, location and difficulty between the source participants.

4.1.2 Data Processing

Data were examined before analyses to qualify for inclusion. In a few instances, seeker participants simply clicked through the survey and did not provide answers to questions. Seeker participant contributions were removed if the timed survey results were under 10 minutes and the participant did not answer any personal data questions.

Personal data guesses were standardized and compared to questionnaire results to determine categorization of answers as well as to preserve source participant privacy. Results were groups into five categories. *Correct* answers provide complete, correct answers to the question as it pertains to each personal datum. For example, in identifying *Children's Names* for a source participant with three children correctly identified in the survey and confirmed in the questionnaire, the *Correct* answer would contain the three correct names. *Correct/Incomplete* answers only contain correct data but do not contain complete data. For example, the personal data response may provide only two of three children's names. *Partially Correct/Incomplete* data contains correct data which is incomplete and also includes additional data and also contains incorrect data. For example, it may contain the names of children but also contain additional names which are not correct. Finally, *Incorrect* answers are completely incorrect guesses. Source participants also recognized that they were not able to find correct answers, often answering "Couldn't find", "impossible" or a similar equivalent. These answers are designated as *Not Found.*

It was necessary to sanitize data for analysis. For personal data guesses, answers were standardized. Capitalization was ignored when considering answers which were

otherwise identical. Additional comments were often supplied by seeker participants. For example, with answers such as “Pete” and “his nickname is Pete”, and “everyone calls him Pete”, respondents all refer to the same nickname. If the actual nickname supplied in the answer is the same, the answers are categorized appropriately as previously described.

Time data required minimal data cleaning. Data was often reported as requested (time in minutes) but was also annotated at times. For example, in one instance, a participant reported “0” minutes and added a note that they had found the answer when searching for a previous data point. These types of exceptions are noted as they occur. Otherwise, times were simply standardized to include only numeric data for analysis. For example, “22 min” or “22 minutes” was recorded as “22”. The time between survey beginning and survey submission was also recorded by the survey software. While this time does not necessarily represent the time on task, it does provide a meaningful comparison with reported times in some instances. Survey time is reported and compared as applicable. In several instances, the time from beginning to end of the survey spanned multiple days. These instances would drastically impact calculations of averages and are unlikely to represent real time on task and were therefore excluded from time calculations.

Location data varied in reporting style. Some participants copied URLs while others reported web locations in a variety of styles including simply listing a particular social media platform or service such as, “Facebook” or “Her facebook page”. When URLs appear, they are listed by the name of the company in order to protect the privacy of the source participants, for example, a Facebook URL referring to a particular source participant would be reported as, “Facebook”. Some participants reported particular search engines as “locations”. These are included in the list as applicable. Additionally,

participants occasionally noted that they searched non-specific locations such as “checked everywhere”. These are reported in the corresponding section as they appear in the data.

Location data was analyzed into general location categories which include social media, government, employer, and commercial. Each category is described as follows. When reporting results, successful answers are provided categorically as well as with specific reference to the particular product, service or organization which provided accurate data.

“Social media venues include weblogs as well as Facebook. Twitter, Foursquare, LinkedIn, Yelp, Flickr, Wikipedia, and Youtube” (Oravec 2012, p. 95). It should be noted that not all social media was self-propagated. At times, social media searches included references to social media accounts of friends or families. Additionally, as social media are increasingly used in business transactions (Vickery, 2015), individual personal information present may not be posted by the person to whom it pertains or it may be posted in the course of business endeavors.

Government websites describe those websites under control of a government entity. In this case, due to the location of the study participants, all government addresses were located within the United States and were designated as .gov addresses. These venues might include things such as searchable state databases.

Employers describe those entities believed, from examining the Uniform Resource Locators (URLS) provided by seeker participants, to belong to employers of the source participant. Employment may or may not be current for source participants. This category included entities such as public school districts, colleges and universities, among others.

Commercial information sources describe information sources that operate as a business without regard to the business model. Seeker participants were instructed only to use free sources and occasionally noted that they were able to obtain partial information and full information required payment. At other times, seeker participants were able to obtain information from commercial sources that do not charge information seekers, at least initially. Commercial sources comprise a variety of commercial endeavors such as commercial enterprises concerned with providing contact information in a phone book style format, commercial sources that support family ancestry research and newspapers.

Finally, to protect the privacy of the source participant, no specific URLs or location designations that would reveal the identity of the source participant are provided in the data analysis. Never the less, the summary data does reveal the locations which are points of vulnerability for personal data on the World Wide Web.

4.1.3 Analytical Process

Three research questions form the crux of the dissertation study. Each pertains to a different aspect of the discoverability of personal data on the World Wide Web, particularly as it pertains to discovery by strangers. Discoverability was explored on the basis of accuracy, location, and difficulty.

To understand the accuracy of personal data, six personal data points, selected as described in Chapter 3, were analyzed. In each case, the source participant was asked to verify the answers provided by seeker participants. The categorization into accuracy groups followed. This analytical approach allowed for a finer grained discernment of accuracy than originally planned but more accurately reflect the result obtained. Accuracy was then

compared by personal data question across the four seeker participants to understand the relative difficulty of retrieving any one personal data point.

The locations of data were examined, first by counting and cataloguing their appearance in reported seeker data and finally by examining locations sorted by accuracy groups. This process allows for comparison of accurate locations to inaccurate search locations. The resulting accurate search locations were examined for trends for each source participant and across all four source participants.

To address difficulty, each personal data point was individually examined with comparison to time metrics as well as to the difficulty scale. ANOVA analysis was performed to differentiate accuracy groups on the variables of self-reported familiarity with internet search, the reported search time and the reported difficulty on the Likert scale. Difficulty may also be observed in the accuracy of each data point, particularly in comparison to time and difficulty variables. Where accuracy and difficulty scores are high and time scores are low, the personal data points are believed to be easier to locate. When longer times are reported and accuracy and difficulty are low, the data points are considered more difficult. Difficulty was compared both internally by source participant and across the entire study.

The analytical approach describes yielded excellent and interesting results that address the research questions by providing insight into the availability of personal data, location of the data and difficulty of accurately retrieving the data.

4.2 Results by Source Participant

Individual results for each source participant are presented. Each source participant is described demographically, followed by a summary and comparative analysis across the seeker participant results pertaining to the source participant. Finally, an in-depth analysis

of each source participants' results, ordered by personal data point are provided. This pattern is repeated four times, one time for each source participant. To protect their privacy, source participants are referred to with initials as DK, GC, KT and OV.

4.2.1 DK Personal Data Seeker Analyses

The following analysis describes the results pertaining to source participant DK. The personal demographics of the source participant are detailed followed by a summary and comparative analysis of each area of measurement in the survey including *familiarity, accuracy of personal data, time on task, location of data,* and *perceived difficulty*. Finally, an in-depth analysis of each personal data point is provided.

4.2.1.1 Demographics and Personal Data Description for DK

Participant DK is a 58-year-old female. Her first name was ranked as most popular in the last 100 years in the early 1980s representing 0.043 percent of female births at that time in the United States (Socail Security Administration, n.d.). Her last name is ranked between 19,000 and 20,000 of 160,975 rankings in the 2010 US Census according to the US Census Bureau (2010).

According to DK's verification questionnaire, her mother's maiden name was not correctly identified. Her nickname was correctly identified. She has three children who were successfully identified and no pets, which was also correctly identified. Her middle name was correctly identified. Her mobile phone number was not identified.

4.2.1.2 Summary and Comparative Analysis for DK

A summary of each data point across the survey pertaining to source participant DK is provided. Comparative analyses of aggregated data are also provided as appropriate.

More detailed description of the results describing each individual data point is provided in section 4.2.2.3.

4.2.1.2.1 Familiarity with Search for DK

Seeker participants were also asked to rate their own familiarity with online search on a five point Likert scale described in methods. The scale ranges from (1) Not at all familiar with conducting online searches for information to (5) Not at all familiar with conducting online searches for information. On average, the participants rated themselves at 3.19 for this source participant. Table 7 provides details about the number and percentage of participants selecting each of the Likert scale questions.

Table 7 Frequency of Familiarity for DK

	Number	**Percentage**
1	6	10.2
2	15	25.4
3	9	15.3
4	20	33.9
5	9	15.3
Total	59	100.0

4.2.1.2.2 Accuracy of Personal Data for DK

Across a total of 59 search participants, no participants correctly identified 4, 5 or 6 data points. Not all data were completely or partially identified by the whole group. Each data point differed in the frequency, time and perceived difficulty. Table 8 summarizes the types of answers provided by the participants that fall into a particular category. For example, 22 supplied participants supplied 0 *Correct* answers, 51 seeker participants supplied 0 *Correct/Incomplete* answers and so forth.

Table 8 Accuracy by Number of Participant Answers for DK

	Correct Answers	Correct/ Incomplete	Partially Correct/ Incomplete	Not Found	Incorrect
0 Answers	22	51	55	15	6
1 Answers	30	8	4	12	15
2 Answers	7	0	0	11	11
3 Answers	0	0	0	13	6
4 Answers	0	0	0	6	6
5 Answers	0	0	0	2	8
6 Answers	0	0	0	0	7

Table 9 describes the accuracy of personal data identified across the total survey. Each data point is described in terms of the number all seeker participants whose answers fall into the various categories. More complete frequency analysis and descriptive statistics for each personal data point are provided in Section 4.2.1.3.

Table 9 Accuracy by Data point for DK

	Correct	Correct/ Incomplete	Partially Correct/ Incomplete	Incorrect	Not found
Mother's Maiden Name	0	0	0	31	28
Nickname	4	0	0	21	31
Children	0	6	4	22	24
Pets	20	0	0	0	28
Middle Name	21	2	0	14	18
Mobile Phone	0	0	0	20	32

4.2.1.2.3 Time on search for DK

Seeker participants reported an average of 103.79 minutes across six personal data points. Analysis of total time on survey reveals actual time on task as an average of 74.45

minutes for 59 participants. For individual questions, participants reported searching for between 1 and 120 minutes. A summary of statistical results is found in Table 1Table 10. Reported time on each individual data point varied widely and will be discussed in the following section.

Table 10 Search Time by Personal Data Point for DK

	N	Range	Min	Max	Mean	SD	Variance
Mother's Maiden Name	55	115	5	120	29.22	3.081	522.174
Nickname	53	63	2	65	18.00	1.807	173.154
Children's Names	56	62	3	65	17.61	1.874	196.606
Pet's Names	49	59	1	60	14.78	1.891	175.136
Middle Name	56	64	1	65	15.70	2.109	249.015
Mobile Phone Number	52	63	2	65	18.62	2.249	263.026

4.2.1.2.4 Location of search for DK

Locations of data, regarding the comparison for accuracy, across the survey varied widely. Some sources were used much more frequently than others. Table 11 provides a summary of data locations. Non-specific data locations, “guessed” or “multiple sources” are described in the results for each personal data point.

Table 11 Location of Personal Data for DK

Locations	Number
411	2
Ancestry	4
AOL	1
Baltimore County Public School	13
Been Verified	6
Birth Records	1
Bing	2
Biography	1
Cityfreq	1
Funeral home guestbook	2
Facebook	44
Google	25

Instant Checkmate	1
LinkedIn	9
Montgomery County Races	1
My Heritage	6
My Life	3
Obituary	3
Pipl	7
Phone book lookup (unspecified)	8
Reverse Phone lookup (unspecified)	1
Social Media (unspecified)	3
Truth Finder	5
Twitter	2
White Pages	35
Elementary School	4
Yahoo	1
Zoom Info	1

4.2.1.2.5 Difficulty of Personal Data Discovery for DK

On average across the survey, participants rated the difficulty as 1.76. This is the average of the six point Likert scale used to describe difficulty of locating data with 1 as "Impossible" and 6 as "Very Easy". The difficulty for each data point is described in Table 12.

Table 12 Personal Data points by Difficulty Scale for DK

	N	Range	Min	Max	Mean	SD	Variance
Mother's Maiden Name	57	4	1	5	1.65	.138	1.089
Nickname	54	4	1	5	1.52	.129	.896
Children's Names	57	4	1	5	1.93	.173	1.709
Pet's Names	50	5	1	6	1.54	.179	1.600
Middle Name	56	5	1	6	2.43	.211	2.504
Mobile Phone Number	57	5	1	6	1.56	.146	1.215

Perceptions of difficulty varied widely among seeker participants. *Nickname* is perceived as the most difficult and *Middle Name* as the easiest. The difficulty scale, with the exception of *Middle Name,* were all relatively close in mean number.

Table 13 compares perceived difficulty in comparison to accuracy from most to least difficult. *Low Score* describes the lowest accuracy and also the lowest number on the Likert scale, with 1 as "Impossible" on the Likert scale. Of interest from a cumulative perspective is the relative accuracy of difficulty assessments compared to accuracy. While no correct answers were provided for *Mother's Maiden Name* or *Mobile Phone Number*, *Mother's Maiden Name* had more incorrect guesses and was therefore scored lower.

Table 13 Comparison of Accuracy to Perceived Difficulty for DK

	Supplied Data Accuracy	**Perceived Difficulty**
Lowest	Mother's Maiden Name	Nickname
	Mobile Phone Number	Pet's Names
	Children	Mobile Phone Number
	Nickname	Mother's Maiden Name
	Pet's Names	Children's Names
Highest	Middle Name	Middle Name

4.2.1.3 Analysis of Individual Personal Data for DK

An analysis of each individual personal data point follows. The results of accuracy, time, location, and difficulty are described and individually compared for accuracy between successful and unsuccessful searches in the accuracy categories described previously.

4.2.1.3.1 Mother's Maiden Name for DK

4.2.1.3.1.1 Accuracy of Mother's Maiden Name for DK

No seeker participants answered *Mother's Maiden Name* correctly. A total of 31 incorrect guesses were provided with 28 reporting failure to discover a potential guess. Percentages and frequencies are reported in Appendix 10.

4.2.1.3.1.2 Familiarity with Search of Mother's Maiden Name for DK

For participant who answered incorrectly, self-assessed familiarity with internet search was (3.29) *Somewhat familiar*, which is a slightly more confident self-assessment than participants to unable to find the question. Descriptive statistics are available in Appendix 11.

4.2.1.3.1.3 Time of Mother's Maiden Name for DK

Time estimates for searches of *Mother's Maiden Name* varied widely, from 5 minutes to 120 minutes. On average, participants reported a mean average of 29.22 minutes spent searching. Complete results are able in Appendix 12.

4.2.1.3.1.4 Location of Mother's Maiden Name for DK

Mother's Maiden Name was searched for in a variety of locations. Facebook was the most popular search location. Appendix 13 provides a complete list of the search locations across the total survey along with frequency measures for this particular personal data point. No locations provided a correct answer.

4.2.1.3.1.5 Difficulty of Mother's Maiden Name for DK

Across seeker participants, *Mother's Maiden Name* was perceived as quite difficult with a mean average difficulty rating of 1.65 with 1 being *"Impossible"* and 6 being "*Very*

Easy". Appendix 14 described the number and percentages of participants selecting each degree of difficulty. Appendix 15 provides descriptive statistics by accuracy groups.

4.2.1.3.2 Nickname for DK

4.2.1.3.2.1 Accuracy of Nickname for DK

DK's *Nickname* was correctly identified by 4 participants. Included were 21 Incorrect guesses and 31 participants reported not finding the answer. Percentages of correct and accurate answers are provided in Appendix 16.

4.2.1.3.2.2 Familiarity with Search of Nickname for DK

For participants who answered correctly, self-assessed familiarity with internet search ranged from (1) *Not Familiar* to (5) *Extremely Familiar*, with a mean average of 3.19. Incorrect participants were more self-confident than participant who reported not locating an answer. Appendix 17 reports descriptive statistics for familiarity by category.

4.2.1.3.2.3 Time of Nickname for DK

Time estimates for searches of *Nickname* varied from 2 minutes to 65 minutes. Appendix 18 describes search time by accuracy grouping. Notably, successful participants reported lower search times.

4.2.1.3.2.4 Location of Nickname for DK

Nickname for DK was searched for a discovered in a variety of locations. All reported locations are provided in Appendix 19. The remaining participants were unable to located the data and left the answer blank. Facebook was, once again, the most popular search location. When compared to accuracy, correct answers were located via commercial sources including *Google, Newspaper Obituary Notices and Truthfinder.*

4.2.1.3.2.5 Difficulty of Nickname Search for DK

Across seeker participants, *Nickname* was perceived as between *Very Difficult* and *Difficult* with a mean average difficulty rating of 1.52 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 20 describes the number of participants selecting each degree of difficulty. Appendix 21 provides descriptive statistical analysis by accuracy group. Successful participants rated the difficulty as 2.

4.2.1.3.3 Children's Names for DK

The results of searches for DK's Children's names follows. The results of accuracy, time, location, and difficulty are described and individually compared between accuracy groups.

4.2.1.3.3.1 Accuracy of Children's Names for DK

Of the participants who supplied answers, none provided complete, correct answers. Complete accuracy results are provided in Appendix 22.

4.2.1.3.3.2 Familiarity with Search of Children's Names for DK

For participants who answered correctly, self-assessed familiarity with internet search ranged from (1) *Not Familiar* to (5) *Extremely Familiar,* with a mean average of 4.08, which is a more confident self-assessment compared to the general average of 3.77 across all participants. Descriptive statistics for familiarity with search for children's names are supplied in Appendix 23.

4.2.1.3.3.3 Time of Search for Children's Names for DK

Time estimates for searches of *Children's Names* varied from 3 minute to 655 minutes. On average, 55 reporting participants provided a mean average of 17.56 minutes

spent searching. The remaining participants did not supply a time. Search time descriptive statistics are supplied in Appendix 24.

4.2.1.3.3.4 Location of Children's Names for DK

Children's Name is discovered primarily on Facebook. 10 total distinct locations were reported. *Facebook* was the most popular search location. While no answers were correct, partially correct/incomplete and correct/incomplete answers were all derived from social media, specifically *Facebook.* Complete location information is available in Appendix 25.

4.1.3.3.5 Difficulty of Children's Names for DK

Across seeker participants, *Children's Names* was perceived as *Difficult* with a mean average difficulty rating of 1.95 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 26 describes the number of participants selecting each degree of difficulty. Compared to the survey sample, successful participants rated the difficulty as 3.67 or easier compared to other accuracy groups. Comparisons by accuracy group are available in Appendix 27.

4.2.1.3.4 Pet's Names for DK

4.2.1.3.4.1 Accuracy of Pet's Names for DK

A total of 15 participants answered *Pet's Names* correctly. Percentages of correct and accurate answers are provided in Appendix 28. Interestingly, no participant provided an incorrect answer.

4.2.1.3.4.2 Familiarity with Search of Pet's Names for DK

For participants who answered correctly, self-assessed familiarity with internet search ranged from (1) *Not Familiar* to (5) *Extremely Familiar,* with mean average of 3.00,

a more conservative estimate than other respondents. Descriptive statistics are provided in Appendix 29.

4.2.1.3.4.3 Time of Search of Pet's Names for DK

Time estimates for searches of *Pet's Names* varied from 1 minutes to 60 minutes with a mean average of 14.88 minutes. When comparing time to accuracy, those participants correctly guessing the *Pet's Names* search times of 1-45 minutes with an average time of 10.65 minutes. Mean average times were shorter for *Correct* answers than *Not Found* answers. Descriptive statistics are provided in Appendix 30.

4.2.1.3.4.4 Location of Search of Pet's Names for DK

Pet's Name was searched for in a variety of locations. *Facebook* continued to be the most popular search location. Appendix 31 provides a complete list of the search locations across the total survey along with frequency measures for this particular personal data point. Correct answers originated from commercial sources including *Google, Yahoo, Pipl* and *Bing*. Commercial, genealogy focused source, *My Heritage,* also provided correct answers. Finally social media resources including unspecified, "*Social Media", and Facebook* were also used to provide accurate answers.

4.2.1.3.4.5 Difficulty of Pet's Names for DK

Across seeker participants, *Pet's Names* received mean average difficulty rating of 1.56 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 32 describes the number of participants selecting each degree of difficulty. Appendix 33 compares the degree of difficulty across the various answer groupings.

4.2.1.3.5 Middle Name for DK

4.2.1.3.5.1 Accuracy of Middle Name for DK

Middle Name was answered correctly by 35 participants. Percentages of correct and accurate answers are provided in Appendix 34. No participants reported skipping the question.

4.2.1.3.5.2 Familiarity with Search of Middle Name for DK

For participants who answered correctly, self-assessed familiarity with internet search ranged from (1) *Not Familiar* to (5) *Extremely Familiar,* with a mean average of 3.19. Participants correctly answering the question scored themselves more critically than those in the *Correct/Incomplete* category and less critically than the *Incorrect* group. Appendix 35 provides complete descriptive statistics for familiarity.

4.2.1.3.5.3 Time of Search of Middle Name for DK

Time estimates for searches of *Middle Name* varied widely, from 1 minute to 65 minutes. On average participants reported a mean average of 15.76 minutes spent searching. Appendix 36 provides complete results of time by accuracy. *Correct* groups reported the lowest search times.

4.2.1.3.5.4 Location of Middle Name for DK

Middle Name was searched for and discovered in a variety of locations. *White Pages* is the most popular search location. Appendix 37 provides a list of the search locations across the total survey along with frequency measures for this particular personal data point. Correct answers were located using commercial venues including *White Pages* and *411.* An employer source, *BCPS,* also yielded correct data.

4.2.1.3.5.5 Difficulty of Middle Name for DK

Across seeker participants, *Middle Name* was perceived as *"Difficult"* with an average difficulty rating of 2.43 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 38 describes the number of participants selecting each degree of difficulty.

Compared to the survey sample, successful participants rated the difficulty as easier compared to unsuccessful participants. Appendix 39 provides descriptive statistics by accuracy.

4.2.1.3.6 Mobile Phone Number for DK

4.2.1.3.6.1 Accuracy of Mobile Phone Number for DK

Mobile Phone Number proved to be a difficult personal data point. No participants answered *Mobile Phone Number* correctly. Appendix 40. provides a summary of accuracy.

4.2.1.3.6.2 Familiarity with Search of Mobile Phone Number for DK

No participants who answered correctly. *Incorrect* participants were more self-critical than *Not Found* participants. Descriptive analytics describing familiarity in terms of search accuracy can be found in Appendix 41.

4.2.1.3.6.3 Time of Mobile Phone Number for DK

Time estimates for searches of *Mobile Phone Number* varied from 2 minutes to 65 minutes. There were no correct times. Mean average search time was 18.62 minutes. Descriptive statistics are provided in Appendix 42.

4.2.1.3.6.4 Location of Mobile Phone Number for DK

Mobile Phone Number was searched for in a variety of locations. An employer, *BCPS* was the most popular search location, followed by commercial venues including

unspecified *"Phone Book"* and *"Phone Book Sites"* and *White Pages.* No correct answer was located. Complete location information is available in Appendix 43.

4.2.1.3.6.5 Difficulty of Mobile Phone Number for DK

Across seeker participants, *Mobile Phone Number* was perceived as quite difficult with a mean average difficulty rating of 1.62 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 44 describes the number of participants selecting each degree of difficulty. Appendix 45 compares perceived difficulty by accuracy group.

4.2.1.3.7 ANOVA for DK

Due to relatively small number of *Correct* groups in several personal data categories, the ANOVA analysis does not consistently represent a statistically significant finding. In several instances there were no *Correct* groups. Analysis of ANOVAs primarily revealed statistically significant differences in difficulty scores when comparing accuracy groups. ANOVAs for DK may be found in Appendices 46-51.

4.2.2 GC Personal Data Seeker Analyses

The following analysis describes the results for source participant GC. The personal demographics of the source participant are detailed followed by a summary and comparative analysis of each area of measurement in the survey including *familiarity, accuracy of personal data, time on task, location of data,* and *perceived difficulty.* Finally, an in-depth analysis of each personal data point is provided.

4.2.2.1 Demographics and Personal Data Description for GC

Participant GC is a 26-year-old female. Her first name ranked in the top five names for girls according to the Social Security Administration at various times between 1990 and 2010 (Social Security Adminsitration, n.d.). Her name also appears as a top 100 names

for births in the last 100 years. Her last name is ranked above 2000 of 160975 rankings in the 2010 US Census according to the United States Census Bureau (2010).

According to HR's verification questionnaire, her mother's maiden name was correctly identified and her nickname were correctly identified. She has three children, two of whom were correctly identified and no pets, which was also correctly identified. She also noted in responding to her questionnaire that seeker participants found the name of a previous pet. Her middle name was correctly identified. Only the last four digits of her mobile phone number were identified.

4.2.2.2 Summary and Comparative Analysis for GC

A summary of each data point across the survey pertaining to source participant GC is provided. Comparative analyses between aggregated data are provided as appropriate. More detailed description of the results in comparing selected areas is provided in Section 4.1.3.

4.2.2.2.1 Familiarity with Search for GC

Seeker participants rate their own familiarity with online search on a five point Likert scale described in methods. The scale ranges from *(1) Not at all familiar with conducting online searches for information* to *(5) Not at all familiar with conducting online searches for information.* On average, the participants rated themselves at 2.94 for this source participant. Table 14 provides details about the number and percentage of participants selecting each of the Likert scale questions.

Table 14 Frequency of Familiarity Selection for GC

	Frequency	**Percent**
1	7	13.7
2	11	21.6

3	17	33.3
4	12	23.5
5	4	7.8
Total	51	100.0

4.2.2.2.2 Accuracy of Personal Data for GC

Across a total of 51 search participants, no participants correctly identified 4, 5 or 6 data points. Each data point was completely or partially identified by seeker participants considering the whole group. Data points differed in the frequency, time, and perceived difficulty. On average, the seeker participants identified 1.67 data points correctly. Table 15 describes the number of participants who found a particular number of answers. For example, 0 seeker participants answered 0 problems correctly.

Table 15 Number of Answers by Category and Seeker Participant for GC

	Correct Answers	**Correct/ Incomplete**	**Partially Correct/ Incomplete**	**Not Found**	**Incorrect**
0 Answers	8	30	37	4	4
1 Answers	32	21	14	18	13
2 Answers	11	0	0	12	19
3 Answers	0	0	0	11	12
4 Answers	0	0	0	5	2
5 Answers	0	0	0	1	1

Table 16 describes the accuracy of personal data identified across the total survey. Each data point is described in terms of the number of all seeker participants falling into a particular correctness category. More complete frequency analysis and descriptive statistics for each personal data point are provided in section 4.2.2.3

Table 16 Accuracy by Data point for GC

	Correct	Correct/ Incomplete	Partially Correct/ Incomplete	Incorrect	Not found
Mother's Maiden Name	1	0	0	31	13
Nickname	3	0	0	21	22
Children	0	4	14	7	8
Pets	15	0	0	24	9
Middle Name	35	0	0	6	8
Mobile Phone	0	1	0	7	34

4.2.2.2.3 Time on search for GC

Seeker participants reported an average of 66.79 minutes across six personal data points. Analysis of total time on survey reveals actual time on task as an average of 45.63 minutes for 49 participants. Two participants were excluded per data cleansing described in Section 4.1.2. For individual questions, participants reported spending as little as 0 minutes to find answers to personal datum questions and searching for as long as 115 minutes. A summary of statistical results is found in Table 17. Reported time on each individual data point varied widely and will be discussed in more detail in each data point section.

Table 17 Search Time by Personal Data point for GC

	N	Min	Max	Mean	SD
Mother's Maiden Name	47	1	115	22.11	20.795
Nickname	49	1	60	14.96	12.721
Children's Names	49	1	60	13.63	12.122
Pet's Names	50	0	60	11.50	12.012
Middle Name	50	1	60	8.50	10.649
Mobile Phone Number	4	5	12	8.25	3.304

4.2.2.2.4 Location of search for GC

Location of data, particularly with regard to comparison for accuracy, across the survey varied widely however, some sources were used much more frequently than others. Table 18 provides a summary of data locations. Some participants reported particular search engines as "locations". These are included in the list as applicable. Additionally, participants occasionally noted that they searched non-specific locations such as "checked everywhere".

Table 18 Location of Personal Data for GC

Location Total Reported	Number
Ancestry	2
AOL	2
BeenVerified	3
Birth Records	2
Facebook	113
GC's Dressage Website	2
Google	19
mn.gov	1
mylife	3
Nuwber	13
Ohio Resident DB	17
Ohio Voters	2
Peoplesmart	1
Phone	2
Quanki	1
Spokeo	1
Truthfinder	11
Twitter	2
VoterRecords	2
White Pages	14

4.2.2.2.5 Difficulty of Personal Data Discovery for GC

On average across the survey, participants rated the difficulty as 2.66. This is the average of the six point Likert scale used to describe difficulty of locating data. The difficulty for each data point is described in Table 19.

Table 19 Personal Data points by Difficulty Scale for GC

	N	Min	Max	Mean	SD
Mother's Maiden Name	47	1	5	2.4	1.393
Nickname	49	1	6	2.14	1.486
Children's Names	49	1	6	3.55	1.608
Pet's Names	50	1	6	3.02	1.708
Middle Name	51	1	6	4.06	1.678
Mobile Phone	50	1	6	1.62	1.338

Perceptions of difficulty varied widely. *Mobile Phone Number* is perceived as the most difficult and *Middle Names* as the easiest. The perception of difficulty at both ends of the scale was matched the accuracy of information provided by participants.

Table 20 compares perceived difficulty in comparison to accuracy from most to least difficult. *Low Score* describes the lowest accuracy and also the lowest number on the Likert scale, with 1 as "Impossible" on the Likert scale. Of interest from a cumulative perspective is the relative accuracy of difficulty assessments compared to accuracy.

Table 20 Comparison of Accuracy to Perceived Difficulty for GC

	Supplied Accuracy	Perceived Difficulty
Lowest	Mobile Phone Number	Mobile Phone Number
	Children's Names	Nickname
	Mother's Maiden Name	Mother's Maiden Name
	Nickname	Pet's Names
	Pet's Names	Children's Names
Highest	Middle Name	Middle Name

4.2.2.3 Analysis of Individual Personal Data points for GC

An analysis of each individual personal data point follows. The results of accuracy, time, location, and difficulty are described and individually compared for accuracy between successful and unsuccessful searches in in degrees as previously described. 47 of 51 participants attempted to answer the question.

4.2.2.3.1 Mother's Maiden Name for GC

4.2.2.3.1.1 Accuracy of Mother's Maiden Name for GC

A single seeker participant answered *Mother's Maiden Name* correctly. A total of 32 incorrect guesses were provided with seventeen reporting searching but failing to discover a potential guess. Percentages of correct and accurate answers are provided in Appendix 52 and compared to incorrect answers.

4.2.2.3.1.2 Familiarity with Search of Mother's Maiden Name for GC

For participant who answered correctly, self-assessed familiarity with internet search was (3) *Somewhat familiar*, which is a the same as the mean of 3 across participants who attempted the question. Participants who believe that they did not find the answer rated themselves slightly lower. Complete descriptive statistics are available in Appendix 53.

4.2.2.3.1.3 Time of Mother's Maiden Name for GC

Time estimates for searches of *Mother's Maiden Name* varied widely, from 1 minutes to 115 minutes. On average, participants reported a mean average of 22.27 minutes spent searching. When comparing time to accuracy, the participant correctly guessing the *Mother's Maiden Name* reported a search time of 15 minutes. Complete results are able in Appendix 54.

4.2.2.3.1.4 Location of Mother's Maiden Name for GC

Mother's Maiden Name was searched for in a variety of locations. Facebook was the most popular search location. Appendix 55 provides a complete list of the search locations across the total survey along with frequency measures for this particular personal data point. When compared to accuracy, the correct result originated from a government source, the *Minnesota Official Marriage System.*

4.2.2.3.1.5 Difficulty of Mother's Maiden Name for GC

Across seeker participants, *Mother's Maiden Name* was perceived as quite difficult with a mean average difficulty rating of 2.40 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 56 describes the number and percentages of participants selecting each degree of difficulty. Compared to the survey sample, successful participants rated the difficulty as 4 or *"Neutral"*, perceiving the question easier compared to the survey as a whole. No participants reported the question as 6, or *"Very Easy".* Appendix 57 compares describes perceived difficulty compared to accuracy groups.

4.2.2.3.2 Nickname for GC

4.2.2.3.2.1 Accuracy of Nickname for GC

GC's *Nickname* was correctly identified by 4 participants. Included were 21 *incorrect* guesses and 24 participants unable to guess. Percentages of correct and accurate answers are provided in Appendix 58.

4.2.2.3.2.2 Familiarity of Nickname for GC

For participants who answered correctly, self-assessed familiarity with search ranged from (1) *Not Familiar* to (5) *Extremely Familiar*, with a mean average of 4.25, which is a more confident self-assessment compared to the general average of 2.96 across all participants. Incorrect participants were also less self-confident than participants who reported not locating an answer. Appendix 59 provides descriptive statistics for familiarity by accuracy category. Two participants did not reply.

4.2.2.3.2.3 Search Time of Nickname for GC

Time estimates for searches of *Nickname* varied from 1 minutes to 60 minutes. Appendix 60 describes search time by accuracy grouping. Notably, successful participants reported lower search times.

4.2.2.3.2.4 Location of Nickname for GC

GC's *Nickname* was searched for a discovered in a variety of locations. All reported locations are provided in Appendix 61. The remaining participants were unable to located an answer and left the survey blank. Facebook was, once again, the most popular search location. When compared to accuracy, all correct answers were located from a social media venue, specifically *Facebook.*

4.2.2.3.2.5 Difficulty of Nickname Search for GC

Across seeker participants, *Nickname* was perceived as between *Very Difficult* and *Difficult* with a mean average difficulty rating of 2.14 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 62 describes the number of participants selecting each degree of difficulty. Compared to the survey sample, successful participants rated the difficulty as 3, perceiving the question as easier compared to the survey as a whole but as more difficult than incorrect participants. Appendix 63 describes the difficulty perception of each accuracy category.

4.2.2.3.3 Children's Names for GC

The results of searches for GC's *Children's Names* follows. The results of accuracy, time, location, and difficulty are described and individually compared for accuracy between successful and unsuccessful searches.

4.2.2.3.3.1 Accuracy of Children's Names for GC

Of the participants who supplied answers, none provided complete, correct answers. One participants reported skipping the question. Complete accuracy results are provided in Appendix 64.

4.2.2.3.3.2 Familiarity with Search of Children's Names for GC

For participants who answered correctly, self-assessed familiarity with internet search ranged from (1) *Not Familiar* to (5) *Extremely Familiar*, with a mean average of 4.08, which is a more confident self-assessment compared to the general average of 3.77 across all participants. Descriptive statistics are provided in Appendix 65.

4.2.2.3.3.3 Time of Search for Children's Names for GC

Time estimates for searches of *Children's Names* varied from 1 minute to 60 minutes. On average, 49 reporting participants provided a mean average of 13.63 minutes spent searching. Two participants did not supply a time. Descriptive statistics are provided in Appendix 66.

4.2.2.3.3.4 Location of Children's Names for GC

Children's Name was searched for and discovered primarily on Facebook. 10 total distinct locations were reported. *Facebook* was the most popular search location. While no answers were correct, partially correct/incomplete and correct/incomplete answers were all derived from the social media venue Facebook. A complete list of search locations is available in Appendix 67.

4.2.2.3.3.5 Difficulty of Children's Names for GC

Across seeker participants, *Children's Names* was perceived as *Difficult* with a mean average difficulty rating of 3.55 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 68 describes the number of participants selecting each degree of difficulty. Compared to the survey sample, successful participants rated the difficulty as slightly easier compared to the survey as a whole. Frequency statistics are provided in Appendix 68. Descriptive statistics are provided in Appendix 69.

4.2.2.3.4 Pet's Names for GC

4.2.2.3.4.1 Accuracy of Pet's Names for GC

A total of 15 participants answered *Pet's Names* correctly. Percentages of correct and accurate answers are provided in Appendix 70. 2 participants skipped the question. Interestingly, the GC noted on her survey that, while currently incorrect, seeker participants had found the name of a pet she previously owned.

4.2.2.3.4.2 Familiarity with Search of Pet's Names for GC

For participants who answered correctly, self-assessed familiarity with internet search ranged from (1) *Not Familiar* to (5) *Extremely Familiar,* with mean average of 2.73, a more conservative estimate than other respondents compared to 2.88 total. Incorrect participants were more confident that participants who reported not finding an answer. Appendix 71 compares familiarity with accuracy of answers.

4.2.2.3.4.3 Time of Pet's Names for GC

Time estimates for searches of *Pet's Names* varied from 0 minutes to 60 minutes with a mean average of 11.55 minutes. When comparing time to accuracy, those participants correctly guessing the *Pet's Names* search times of 0-60 minutes with an average time of 14.60 minutes. Mean average times were shorter for correct answers than incorrect or not found answers. Appendix 72 provides descriptive statistics.

4.2.2.3.4.4 Location of Pet's Names for GC

Pet's Name was searched for in a variety of locations. *Facebook* continued to be the most popular search location. Appendix 73 provides a complete list of the search locations across the total survey along with frequency measures for this particular personal

data point. Correct answers originated from commercial sources, namely *Google* and *Truthfinder,* and from the social media venue *Facebook*.

4.2.2.3.4.5 Difficulty of Pet's Names for GC

Across seeker participants, *Pet's Names* received mean average difficulty rating of 3.06 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 74 described the number of participants selecting each degree of difficulty. Appendix 75 compares the degree of difficulty across the various answer groupings. Interestingly, incorrect answers perceived the question is significantly easier than correct and not found groups.

4.2.2.3.5 Middle Name for GC

4.2.2.3.5.1 Accuracy of Middle Name for GC

Middle Name was answered correctly by 35 participants. Percentages of correct and accurate answers are provided in Appendix 76. No participants reported skipping the question.

4.2.2.3.5.2 Familiarity with Search of Nickname for GC

For participants who answered correctly, self-assessed familiarity with internet search ranged from (1) *Not Familiar* to (5) *Extremely Familiar,* with a mean average of 2.92. Participants correctly answering the question scored themselves more critically than those in the incorrect category and less critically than those reporting that they did not find the answer. Appendix 77 provides descriptive statistics for familiarity by accuracy group.

4.2.2.3.5.3 Time of Middle Name for GC

Time estimates for searches of *Middle Name* varied widely, from 1 minute to 60 minutes. On average participants reported a mean average of 8.47 minutes spent searching. Appendix 78 provides complete results of time by accuracy.

4.2.2.3.5.4 Location of Middle Name for GC

Middle Name was searched for and discovered in a variety of locations. The *Ohio Resident Database* is the most popular search location, followed by *Nuwber*. Appendix 79 provides a list of the search locations across the total survey along with frequency measures for this particular personal data point. Correct answers were located on a government site, *Ohio Resident Database.* Commercial sites *Nuwber, Been Verified, Google, White Pages, MyLife, Truthfinder, VoterRecords, Peoplesmart, Whitepages, Quanki,* and *Instant Checkmate* also provided accurate answers.

4.2.2.3.5.5 Difficulty of Middle Name for GC

Across seeker participants, *Middle Name* was perceived as *"Difficult"* with an average difficulty rating of 4.06 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 80 describes the number of participants selecting each degree of difficulty. Compared to the survey sample, successful participants rated the difficulty as easier compared to unsuccessful participants. Appendix 81 provides descriptive statistics by accuracy.

4.2.2.3.6 Mobile Phone Number for GC

4.2.2.3.6.1 Accuracy of Mobile Phone Number for GC

Mobile Phone Number proved to be a difficult personal data point. No participants answered *Mobile Phone Number* correctly. One participant successfully identified the last four digits. Appendix 82 provides a summary of accuracy.

4.2.2.3.6.2 Familiarity with Search of Mobile Phone Number for GC

As no participants answered correctly, familiarity with search cannot be evaluated for comparison. For the participants who provided partially correct answer, self-assessed

familiarity was rated at 2 on a scale of (1) *Not Familiar* to (5) *Extremely Familiar*. The more successful participant was more self-critical than unsuccessful participants. Descriptive statistics are provided in Appendix 83.

4.2.2.3.6.3 Time of Mobile Phone Number for GC.

Time estimates for searches of *Mobile Phone Number* varied from 5 minutes to 12 minutes. There were no correct times, the Partially Correct answer did not supply a time. Only four participants supplied a time. For those participants supplying a time, the mean time was 8.25 minutes.

4.2.2.3.6.4 Location of Mobile Phone Number for GC

Mobile Phone Number was searched for in a variety of locations. Ten total distinct locations were reported. *Facebook* remained the most popular search location, followed by *WhitePages* and *Google.* Partially Correct/Incomplete answer was found via commercial source, *Truthfinder.* The complete list of search locations is provided in Appendix 84.

4.2.2.3.6.5 Difficulty of Mobile Phone Number for GC

Across seeker participants, *Mobile Phone Number* was perceived as quite difficult with a mean average difficulty rating of 1.62 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 85 described the number of participants selecting each degree of difficulty. Appendix 86 provides descriptive statistics of difficulty by accuracy group.

4.2.2.3.7 GC ANOVA

Due to relatively small number of *Correct* groups in several personal data categories, the ANOVA analysis does not consistently represent a statistically significant finding. Analysis of ANOVAs primarily revealed statistically significant differences in

difficulty scores when comparing accuracy groups. ANOVAs for GC may be found in Appendices 87-92.

4.2.3 KT Personal Data Seeker Analyses

The following analysis describes the results for source participant KT. The personal demographics of the source participant are detailed followed by a summary and comparative analysis of each area of measurement in the survey including *familiarity, accuracy of personal data, time on task, location of data,* and *perceived difficulty.* Finally, an in-depth analysis of each personal data point is provided.

4.2.3.1 Demographics and Personal Data Description for KT

Participant KT is a 65-year-old male. His first name has not ranked in the 100 top names in the last 100 years. It was most popular when, in the early 1940s it represented 0.039% of all male births one year according to the Social Security Administration (Administration, n.d.). His first name is also somewhat gender ambiguous, with more female than male occurrences in the last 100 years (Social Security Administration, n.d). His last name is ranked above 200 of 160975 surnames in the 2010 United States Census according to the United States Census Bureau (2010).

After cleansing, 62 seeker participants provided information about KT. According to KT's verification questionnaire, his mother's maiden name was correctly identified and his nickname was correctly identified. He has four children, all of whom were correctly identified and a pet whose name was correctly identified. His middle name was correctly identified as was his mobile phone number.

4.2.3.2 Summary and Comparative Analysis for KT

A summary of each data point across the survey pertaining to source participant KT is provided. Comparative analysis between the aggerate data are provided as appropriate. More detailed description of the results in comparing selected areas is provided in Section 4.1.3.

4.2.3.2.1 Familiarity with Search for KT

Prior to search, seeker participants rated their own familiarity with online search on a five point Likert scale described in Chapter 3. The scale ranges from (1) Not at all familiar with conducting online searches for information to (5) Extremely familiar with conducting online searches for information. On average, the participants rated themselves at 3.15 for this source participant. Table 21 provides details about the number and percentage of participants selecting each of the Likert scale questions.

Table 21 Frequency of Familiarity for KT

	Frequency	**Percent**
1	6	9.7
2	10	16.1
3	23	37.1
4	14	22.6
5	9	14.5
Total	62	100.0

4.2.3.2.2 Accuracy of Personal Data for KT

Across a total of 62 search participants, no participants correctly identified more than 3 data points. Each point was identified by the whole group. Each data point differed in the frequency, time and perceived difficulty. Table 22 summarizes the types of answers provided by the participants that fall into a particular category. In this instance, 29

participants answered 1 question correctly and only 2 participants correctly identified 3 answers.

Table 22 Accuracy by Number of Participant Answers for KT

	Correct Answers	Correct/ Incomplete	Partially Correct/ Incomplete	Not Found	Incorrect
0 Answers	22	43	61	10	16
1 Answers	29	14	0	15	15
2 Answers	8	4	0	14	10
3 Answers	2	0	0	9	14
4 Answers	0	0	0	7	4
5 Answers	0	0	0	3	1
6 Answers	0	0	0	3	1

Table 23 describes the accuracy of personal data identified across the survey. Each data point is described in terms of the number of participants correctly identifying the particular datum. More complete frequency analysis and descriptive statistics for each personal data point are provided in Section 4.2.3.3.

Table 23 Accuracy by Data point for KT

	Correct	Correct/ Incomplete	Partially Correct/ Incomplete	Incorrect	Not found
Mother's Maiden Name	2	0	0	19	38
Nickname	10	0	3	25	13
Children Names	1	13	21	11	3
Pets Names	1	0	0	25	9
Middle Name	33	9	0	10	4
Mobile Phone	4	0	0	41	5

4.2.3.2.3 Time on search for KT

Seeker participants reported an average of 83.09 minutes across six personal data points. Analysis of total time on survey as reported by the survey software reveals actual time on task as an average of 44.16 minutes for 57 participants. The remaining

participants were excluded from the clock time calculations as described previously. For individual questions, participants reported spending as little as 0 minutes to find data and searching for as long as 115 minutes. In one instance, the 0 minute was annotated by the survey taker saying, “found when searching for another question”. The participant did not specify which search assisted them in identifying multiple personal data points. A summary of statistical results is found in Table 24 Reported time on each individual data point varied widely and will be discussed in the following section.

Table 24 Search Time by Personal Data point for KT

	N	Min	Max	Mean	SE	SD	Variance
Mother's Maiden Name	59	1	120	22.78	2.579	19.807	392.313
Nickname	56	3	45	14.34	1.374	10.284	105.756
Children's Names	57	0	60	14.88	1.648	12.441	154.788
Pet's Names	44	2	60	16.25	1.893	12.557	157.680
Middle Name	57	0	50	11.95	1.529	11.547	133.336
Mobile Phone Number	53	1	60	12.79	1.706	12.416	154.168

4.2.3.2.4 Location of search for KT

Locations of data, particularly with regard to comparison for accuracy, across the survey varied widely however, some sources were used much more frequently than others. Table 25 provides a summary of data locations. Some participants reported particular search engines as “locations”. These are included in the list as applicable. Additionally, participants occasionally noted that they searched non-specific locations such as “everywhere” or “social media”.

Table 25 Location of Personal Data for KT

Locations	Number
Ancestry	1
BeenVerified	18
California Public records	9
Church Website	1
dbcomp.co	1

Desert Christian School	1
Life Pacific College	22
Family Search	1
Facebook	58
Google	29
Instant Check Mate	1
Intelius	3
Mylife	3
Nuwber	25
Lancaster People	2
Linked In	7
Obituary	5
People Smart	1
Pipl	1
Phone	3
Public Records 360	7
Quanki	4
Radaris	2
Rate My Professor	4
Spokeo	4
Truth Finder	7
Twitter	2
Wikipedia	2
White Pages	30
Yellow Pages	2
Youtube.com	6
Zabasearch	1

4.2.3.2.5 Difficulty of Personal Data Discovery for KT

On average across the survey, participants rated the difficulty as 2.70. This is the average of the six point Likert scale used to describe difficulty of locating each data point. The difficulty for each data point is described in Table 26.

Table 26 Personal Data points by Difficulty Scale for KT

	N	Min	Max	Mean	SE	Variance
Mother's Maiden Name	59	1	5	1.86	.172	1.319
Nickname	56	1	6	2.64	.215	1.612
Children's Names	58	1	6	2.90	.215	1.640
Pet's Names	48	1	6	1.94	.216	1.493
Middle Name	59	1	6	3.59	.210	1.609
Mobile Phone	54	1	6	3.39	.233	1.709

Perceptions of difficulty varied widely. *Mother's Maiden Name* is perceived as the most difficult and *Middle Names* as the easiest. *Mother's Maiden Name* was substantially easier based on successful searches than it was perceived to be. Children's names, for a complete, correct list, was also more difficult than perceived. *Middle Name* was the easiest both from perception and actual correct answers.

Table 27 compares perceived difficulty in comparison to accuracy from most to least difficult. *Low Score* describes the lowest accuracy and also the lowest number on the Likert scale, with 1 as "Impossible" on the Likert scale. Of interest from a cumulative perspective is the relative accuracy of difficulty assessments compared to accuracy.

Table 27 Comparison of Accuracy to Perceived Difficulty for KT

	Supplied Accuracy	Perceived Difficulty
Lowest Score	Pet's Names	Mother's Maiden Name
	Children's Names	Pet's Names
	Mother's Maiden Name	Nickname
	Mobile Phone Number	Children's Names
	Nickname	Mobile Phone Number
Highest Score	Middle Name	Middle Name

4.2.3.3 Analysis of Individual Personal Data points for KT

An analysis of each individual personal data point follows. The results of accuracy, time, location, and difficulty are described and individually compared for accuracy between successful and unsuccessful searches in in degrees as previously described.

4.2.3.3.1 Mother's Maiden Name for KT

4.2.3.3.1.1 Accuracy of Mother's Maiden Name for KT

Three seeker participants answered *Mother's Maiden Name* correctly. A total of 19 incorrect guesses were provided. Percentages of correct and accurate answers are provided in Appendix 93 and compared to incorrect answers.

4.2.3.3.1.2 Familiarity for Mother's Maiden Name for KT

For participant who answered correctly, self-assessed familiarity with internet search was 3.5, which is slightly higher than the mean across all participants. Participants who believe that they did not find the answer rated themselves slightly lower. Complete descriptive statistics are available in Appendix 94.

4.2.3.3.1.3 Time of Mother's Maiden Name for KT

Time estimates for searches of *Mother's Maiden Name* varied widely, from 1 minutes to 120 minutes. On average, participants reported a mean average of 22.78 minutes spent searching. When comparing time to accuracy, the participant correctly guessing the *Mother's Maiden Name* reported a search time of 57.50 minutes, nearly three times as long as unsuccessful searches. Complete results are able in Appendix 95.

4.2.3.3.1.4 Location of Mother's Maiden Name for KT

Mother's Maiden Name was searched for in a variety of locations. *Facebook* was the most popular search location. Appendix 96 provides a complete list of the search locations across the total survey along with frequency measures for this particular personal data point. When compared to accuracy, correct results were obtained from a commercial sources *Newspaper Obituary* and commercial genealogy sources *familysearch.org*.

4.2.3.3.1.5 Difficulty of Discovery of Mother's Maiden Name for KT

Across seeker participants, *Mother's Maiden Name* was perceived as quite difficult with a mean average difficulty rating of 1.86 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 97 describes the number and percentages of participants selecting each degree of difficulty. Compared to the survey sample, successful participants rated the difficulty as 2.50, perceiving the question easier compared to the survey as a whole. No participants reported the question as 6, or *"Very Easy".* Appendix 98 provides descriptive statistics of difficulty by accuracy groups.

4.2.3.3.2 Nickname for KT

4.2.3.3.2.1 Accuracy of Nickname for KT

KT's *Nickname* was correctly identified by 10 participants. Percentages of correct and accurate answers are provided in Appendix 99.

4.2.3.3.2.1 Familiarity of Nickname for KT

For participants who answered correctly, self-assessed familiarity with internet search ranged from (1) *Not Familiar* to (5) *Extremely Familiar,* with a mean average of 3.90, which is a more confident self-assessment compared to the general average of 3.16 across all participants. Incorrect participants were also less self-confident than participant who reported not locating an answer. Appendix 100 reports descriptive statistics for familiarity by category.

4.2.3.3.2 Search Time of Nickname for KT

Time estimates for searches of *Nickname* varied from 3 minutes to 45 minutes. Appendix 101 describes search time by accuracy grouping. Notably, successful

participants reported lower search times than the average but longer search times than *Partially Correct/Incomplete* participants.

4.2.3.3.2 Location of Nickname for KT

Nickname was searched for a discovered in a variety of locations. All reported locations are provided in Appendix 102. Facebook was, once again, the most popular search location. When compared to accuracy, correct answers were located on social media specifically *Facebook and Rate My Professor* and an employer, *Life Pacific College.* Commercial sources *Google,* and *Truthfinder* also yielded correct answers. Interestingly, while *Facebook* was the most searched, it did not provide the most consistent correct answers as 3 of 4 answers obtained on *Rate My Professor* were correct compared to 3 of 14 answers on *Facebook.*

4.2.3.3.2.5 Difficulty of Nickname Search for KT

Across seeker participants, *Nickname* received a mean average difficulty rating of 2.64 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 103 the number of participants selecting each degree of difficulty. Compared to the survey sample, successful participants rated the difficulty as 4.10, perceiving the question as easier compared to the survey as a whole as well as easier than the other participant groups. Appendix 104 describes the difficulty perception of each accuracy category.

4.2.3.3.3 Children's Names for KT

The results of searches for KT's Children's names follows. The results of accuracy, time, location, and difficulty are described and individually compared for accuracy between successful and unsuccessful searches.

4.2.3.3.3.1 Accuracy of Children's Names for KT

Of the participants who supplied answers, one provided a complete, correct answer. Fare more frequent were Partially Correct/ Incomplete answers in which both accurate and inaccurate information was supplied. Complete accuracy results are provided in Appendix 105.

4.2.3.3.3.2 Familiarity with Search of Children's Names for KT

For participants who answered correctly, self-assessed familiarity with internet search ranged from (1) *Not Familiar* to (5) *Extremely Familiar,* with a mean average of 3, a very similarity self-assessment compared to the general average of 3.14 across all participants. Incorrect participants ranked slightly lower than average. Descriptive statistics are provided in Appendix 106.

4.2.3.3.3.3 Time of Search for Children's Names for KT

Time estimates for searches of *Children's Names* varied from 0 minute to 60 minutes. On average, 57 reporting participants provided a mean average of 14.88 minutes spent searching. The remaining participants did not supply a time. Descriptive statistics are provided in Appendix 107.

4.2.3.3.3.4 Location of Children's Names for KT

Children's Name was searched for and partially discovered primarily on social media resrouce *Facebook.* Additionally, some seeker participant provided more insight into the search process by describing, "circular queries using bits of information mined by google search links > [Newspaper website with obituary]" and, "her wife's Facebook page". Some participants provided non-specific results that do not seem to indicate a specific company or website. These included nonspecific *"Public Records", "People search" and "Image search".* The one fully correct answer was provided from a

commercial source *intelius.com.* A complete list of search locations is provided in Appendix 108.

4.2.3.3.3.5 Difficulty of Children's Names for KT

Across seeker participants, *Children's Names* was perceived 2.90 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 109 describes the number of participants selecting each degree of difficulty. Descriptive statistics by accuracy group are provided in Appendix 110. Compared to the survey sample, successful participants rated the difficulty as slightly more difficult than the *Incomplete* groups but as easier than the *Incorrect* and *Not found* groups.

4.2.3.3.4 Pet's Names for KT

4.2.3.3.4.1 Accuracy of Pet's Names for KT

A total of 15 participants answered *Pet's Names* correctly. Percentages of correct and accurate answers are provided in Appendix 111. Only one correct answer was provided.

4.2.3.3.4.2 Familiarity with Search for Pet's Names for KT

For the participant who answered correctly, self-assessed familiarity with internet was rated as 5 - *Extremely Familiar.* This is a significantly more confident self-*assessment* compared to 3.06 mean total. Appendix 112 compares familiarity with accuracy of answers.

4.2.3.3.4.3 Time for Pet's Names for KT

Time estimates for searches of *Pet's Names* varied from 2 minutes to 60 minutes with a mean average of 16.25 minutes. When comparing time to accuracy, reported a search time of 3 minutes. Mean average times were significantly shorter. Descriptive statistics are provided in Appendix 113.

4.2.3.3.4.4 Location of Pet's Names for KT

Pet's Name was searched for in a variety of locations. *Facebook* continued to be the most popular search location. Appendix 114 provides a complete list of the search locations across the total survey along with frequency measures for this particular personal data point. One successful search was conducted with the correct answer provided from social media source *Facebook.*

4.2.3.3.4.5 Difficulty of Pet's Names for KT

Across seeker participants, *Pet's Names* received mean average difficulty rating of 1.94 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 115 describes the number of participants selecting each degree of difficulty. Appendix 116 compares the degree of difficulty across the various answer groupings. The correct group perceived the question as being very easy compared to incorrect groups.

4.2.3.3.5 Middle Name for KT

4.2.3.3.5.1 Accuracy of Middle Name for KT

Middle Name was answered correctly by 33 participants. Percentages of correct and accurate answers are provided in Appendix 117.

4.2.3.3.5.2 Familiarity with Search of Middle Name for KT

For participants who answered correctly, self-assessed familiarity with internet search ranged from (1) *Not Familiar* to (5) *Extremely Familiar,* with a mean average of 3.10. Participants correctly answering the question scored themselves more critically than participants who did not find the answer. Appendix 118 provides complete descriptive statistics for familiarity by accuracy group.

4.2.3.3.5.3 Time of Search of Middle Name for KT

Time estimates for searches of *Middle Name* varied widely, from 0 minute to 50 minutes. On average participants reported a mean average of 11.95 minutes spent searching. Appendix 119 provides complete results of time by accuracy. Correct answer time was above average but shorter than *Incorrect* time.

4.2.3.3.5.4 Location of Middle Name for KT

Middle Name was searched for and discovered in a variety of locations. *Facebook* was the most popular search location, followed by *Nuwber.* Appendix 120 provides a list of the search locations across the total survey along with frequency measures for this particular personal data point. The non-specific "*Public Search"* was also listed. Correct answers were located on a social media source, *Facebook.* Multiple commercial sources also provided correct results including *White Pages, Truthfinder, Youtube, Public Records 360, Nuwber, Mylife, Spokeo, Been Verified, Google, White Pages,* and *Yellow Pages*.

4.2.3.3.5.5 Difficulty of Search for Middle Name for KT

Across seeker participants, *Middle Name* was perceived as *"Difficult"* with an average difficulty rating of 4.06 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 121 describes the number of participants selecting each degree of difficulty.

Compared to the survey sample, successful participants rated the difficulty as easier compared to unsuccessful participants. Descriptive statistics are provided in Appendix 122. provides descriptive statistics by accuracy.

4.2.3.3.6 Mobile Phone Number for KT

4.2.3.3.6.1 Accuracy of Mobile Phone Number for KT

Mobile Phone Number proved to be a difficult personal data point for many seeker participants. Four participants answered *Mobile Phone Number* correctly. Appendix 123 provides a summary of accuracy.

4.2.3.3.6.2 Familiarity with Search of Mobile Phone Number for KT

For participants who answered correctly, self-assessed familiarity with internet search ranged from (1) *Not Familiar* to (5) *Extremely Familiar,* with a mean average of 3.09. The more successful participant was more self-critical than unsuccessful participants. Descriptive statistics for familiarity by accuracy groups are provided in Appendix 124.

4.2.3.3.6.3 Search Time of Mobile Phone Number for KT

Time estimates for searches of *Mobile Phone Number* varied from 1 minutes to 60 minutes. The mean time for correct guesses was 26.5, more than twice the amount of time for *Incorrect* and *Not Found* guesses. Descriptive statistics by accuracy group are provided in Appendix 125.

4.2.3.3.6.4 Location of Search of Mobile Phone Number for KT

Mobile Phone Number was searched for in a variety of locations with *Nuwber* and *Been Verified* as the most popular search locations. Correct answers were successfully located via social networking site *Youtube.* While apparently self-posted via social media, the video was an introductory course lesson taught by the participant in his employment role. The video included his mobile phone number as a contact point for students. A complete list of search locations appears in Appendix 126.

4.2.3.3.6.5 Difficulty of Search of Mobile Phone Number for KT

Across seeker participants, *Mobile Phone Number* was perceived as quite difficult with a mean average difficulty rating of 1.62 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 127 the number of participants selecting each degree of difficulty. Appendix 128 compares perceived difficulty by accuracy group.

4.2.3.3.7 KT ANOVA

Due to relatively small number of *Correct* groups in several personal data categories, the ANOVA analysis does not consistently represent a statistically significant finding. Analysis of ANOVAs primarily revealed statistically significant differences in difficulty scores when comparing accuracy groups. ANOVAs for KT may be found in Appendices 129-134

4.2.4 OV Personal Data Seeker Analyses

The following analysis describes the results for source participant OV. The personal demographics of the source participant are detailed followed by a summary and comparative analysis of each area of measurement in the survey including *familiarity, accuracy of personal data, time on task, location of data,* and *perceived difficulty.* Finally, an in-depth analysis of each personal data point is provided

4.2.4.1 Demographics and Personal Data Description for OV

Participant OV is a 31-year-old male. His surname ranks between 3000 and 4000 of 160975 rankings in the 2010 U.S Census. His first name ranks consistently high in the last 100 years of Social Security Administration records. It was most popular in 1955. In

that year, 11,007 male babies were given the name, representing 0.527 percent of total male births. It consistently ranks in the low 200 or higher in the last 100 years.

According to OV's verification questionnaire, his mother's maiden name was correctly identified and his nickname was correctly identified. He has no children and no pets, which were both correctly identified. His mobile phone number was identified. Seeker participants also provided, unsolicited, OV's address, which was also correctly identified per the source participant. The source participant identified the correct identification of his address on the verification questionnaire with the additional notation which read, "scary" (OV-Verification Questionnaire).

4.2.4.2 Summary and Comparative Analysis for OV

A summary of each data point across the survey pertaining to source participant OV is provided. Comparative analyses between the aggerate data are also provided as appropriate. More detailed description of the results in comparing selected areas is provided in Section 4.2.43.

4.2.4.2.1 Familiarity with Search for OV

Seeker participants provided a ranking of their own familiarity with online search on a five point Likert scale previously described. On average, the participants rated themselves at 3.77 for this source participant. Across the survey, participants tended to view themselves as *"somewhat" or "moderately" familiar with online searches for information*. Table 28 provides details about the percentage of participants selecting each of the Likert scale questions.

Table 28 Familiarity with online searches for information for OV

	Frequency	Percent
1	6	11.3
2	12	22.6
3	15	28.3
4	16	30.2
5	4	7.5
Total	53	100.0

4.2.4.2.2 Accuracy of Personal Data for OV

Across a total of 53 search participants, no participant correctly identified all six data points. Each point was correctly identified collectively by the group. Participants varied in the amount of data they were able to correctly ascertain. Table 29 describes the number of participants who found a particular number of answers. For example, 4 seeker participants answered 0 problems correctly.

Table 29 Number of Answers by Category and Seeker Participant for OV

	Correct	Correct/ Incomplete	Partially Correct/ Incomplete	Incorrect	Not Found
0 Answers	4	37	51	15	11
1 Answer	10	16	2	21	15
2 Answers	16	0	0	12	17
3 Answers	13	0	0	3	4
4 Answers	8	0	0	2	6
5 Answers	2	0	0	0	0
6 Answers	0	0	0	0	0

Each data point differed in the frequency, time and perceived difficulty. On average, the seeker participants identified 2.51 data points correctly. Table 30 describes the accuracy of personal data identified across the total survey. Each data point are described in terms of the total number of guesses by accuracy groups.

Table 30 Accuracy by Data point for OV

	Correct	Correct/ Incomplete	Partially Correct/ Incomplete	Incorrect	Not found
Mother's Maiden Name	4	0	2	16	29
Nickname	15	0	0	25	6
Children's Names	40	0	0	5	6
Pets Names	35	0	0	0	10
Middle Name	24	15	0	3	3
Mobile Phone	5	1	0	7	31

4.2.4.2.3 Time on search for OV

Seeker participants reported an average of 79.08 minutes across six personal data points. Analysis of total time on survey reveals actual time on task as an average of 52.73 minutes. Total times on reported by survey software from beginning of the survey from completion ranged from 9 minutes to 625 minutes. Reported times ranged from 15 minutes to 360 minutes. For individual questions, participants reported spending as little as 1 minute to find a data point and searching for as long as 120 minutes. Reported time on each individual data point varied widely and will be discussed in Section 4.2.4.3

Table 31 Search Time by Personal Data point for OV

	N	***Min***	***Max***	***Mean***	***SE***	***SD***
Mother's Maiden Name	50	5	120	22.48	2.938	20.775
Nickname	48	1	60	17.21	2.171	15.042
Children's Names	51	1	60	14.29	1.644	11.743
Pet's Names	43	2	60	12.86	1.692	11.094
Middle Name	51	1	56	13.65	1.680	11.998
Mobile Phone Number	44	1	120	21.77	3.353	22.243

4.2.4.2.4 Location of search for OV

Locations of datum varied widely however, some sources were used much more frequently than others. Table 32 provides a summary of data locations which appeared

more than once. Additionally, some participants noted “lack of evidence” as support for some data points, particularly *Pet’s Names* and *Children’s Names* to conclude that there were no children or no pets. Additionally, one participant noted an answer as “a guess”. Some participants reported particular search engines as “locations”. These are included in the list as applicable.

Table 32 Location of Personal Data for OV

Locations	**Number**
Academia	1
Ancestry	1
AOL	1
Been Verified	2
Bing	5
Birth Records	2
Blog	7
Educause	1
Facebook	119
Google	15
LinkedIn	10
Nuwber	2
Phone Website (unspecified)	1
Pipl	1
Quanki	3
Research Gate	1
Spokeo	7
Snapchat	1
Twitter	20
University of Washington	6
White Pages	29
Wikipedia	1
Wikitree	1
Yellow Pages	2

4.2.4.2.5 Difficulty of Personal Data Discovery for OV

On average across the survey, participants rated the difficulty as 2.51. This is the average of the six point Likert scale used to describe difficulty of locating data. Of interest from a cumulative perspective is the relative accuracy of difficulty assessments compared to accuracy. Table 33 describes the perceived difficulty compared to the difficulty as described by correct answers. The discrete number of correct guesses is supplied along with the total number of guesses. This does not include the number of participants who described the data point as "impossible" to obtain and did not provide a guess. It should be noted that the number of guesses is the total number of guesses, not the number of unique guesses. The data are described in more detail for each data point in section 4.2.4.3

Table 33 Personal Data points by Difficulty Scale for OV

	N	Min	Max	Mean	SE
Mother's Maiden Name	51	1	5	2.00	.210
Nickname	48	1	6	2.67	.221
Children's Names	52	1	6	3.06	.219
Pet's Names	46	1	5	2.41	.198
Middle Name	51	1	6	3.37	.233
Mobile Phone	45	1	5	1.69	.176

Perceptions of difficulty varied widely. Some data points are perceived as more difficult despite better success. For example, *Pet's Names* are perceived as more difficult than *Nickname*, despite much better accuracy for *Pet's Names. Middle Name* is considered easiest question but ease is not reflected in actual success. For these questions, participants perceived the question as easier than it actually was in reflected in the number of correct guesses.

At times, question difficulty perception may be influenced by the seeker participants' perception of the ability to find an answer. *Mobile Phone Number* was perceived as the most difficult data point and was also the data point which received the least *N* of guesses. Despite the same *N* of correct guesses for *Mother's Maiden Name* and *Mobile Phone Number*, there were a significantly lower number of guesses for *Mobile Phone Number*. However, the percentage of correct guesses of total guesses is significantly higher for *Mobile Phone Number* compared to *Mother's Maiden Name.*

Affecting the *Mobile Phone Number* datum in particular is the possibility of user error. While all the *Mother's Maiden Name*s contain some type of name, the *Mobile Phone Number* contained several data points which are obviously not phone numbers, having the incorrect number of digits for local, US or international phone calls. Each of these incorrect guesses was in the format of a zip code. When these incorrect guesses are removed, correct guesses compared to number of guesses jumps to 55%. In other words, for those participants who provided a phone number format answer, 55% were correct.

While *Mobile phone number* is perceived as more difficult, on most measures, *Mobile Phone Number* out performs *Mother's Maiden Name* on many levels. Given the significantly smaller number of guesses, this may be influenced by the seeker participant's awareness that they did not find an answer.

Table 34 compares perceived difficulty in comparison to accuracy from most to least difficult. *Low Score* describes the lowest accuracy and the lowest number of difficulty assigned with 1 as "Impossible" on the Likert scale.

Table 34 Comparison of Accuracy to Perceived Difficulty for OV

	Supplied Accuracy	Perceived Difficulty
Lowest	Mother's Maiden Name	Mobile Phone Number
	Mobile Phone Number	Mother's Maiden Name
	Nickname	Pet's Names
	Middle Name	Nickname
	Pet's Names	Children's Names
Highest	Children's Names	Middle Name

4.2.4.3 Analysis of Individual Personal Data points for OV

An analysis of each individual personal data point follows. The results of accuracy, time, location, and difficulty are described and individually compared for accuracy between groups.

4.2.4.3.1 Mother's Maiden Name for OV

4.2.4.3.1.1 Accuracy of Mother's Maiden Name for OV

As described in the literature, "Mother's Maiden Name" is often used as an authenticator. Fortunately, it was also one of the more difficult personal data points for this particular participant. Four participants correctly identified the *Mother's Maiden Name* and two included the correct name in a list of names, rendering their answers partially correct. Numbers and percentages by accuracy category are provided in Appendix 135.

4.2.4.3.1.2 Familiarity with Search of Mother's Maiden Name for OV

For participant who answered correctly, self-assessed familiarity with internet search was 2.75, the lowest self-rated groups. Participants who believe that they did not find the answer rated themselves slightly higher. Complete descriptive statistics are available in Appendix 136.

4.2.4.3.1.3 Time of Mother's Maiden Name for OV

Time estimates for searches of *Mother's Maiden Name* varied widely, from 5 minutes to 120 minutes. On average, 50 participants reported a mean average of 22.48 minutes spent searching. Two participants did not attempt the question and one participant did not supply a time.

When comparing time to accuracy, those participants correctly guessing the *Mother's Maiden Name* reported search times of 10-30 minutes with an average time of 17.75 minutes. Descriptive statistics by accuracy group are provided in Appendix 137

4.2.4.3.1.4 Location of Mother's Maiden Name for OV

Mother's Maiden Name was searched for and discovered in a variety of locations. 31 total distinct locations were reported. Facebook was by far the most popular search location, followed by *Google, LinkedIn, University of Washington,* and *WhitePages.* Appendix 138 provides a complete list of the search locations across the total survey along with frequency measures for this particular personal data point. When compared to accuracy, all the correct answers, save one, were located using social media, particularly *Facebook*. The remaining correct answer was located using commercial genealogy site *Moose-Roots Birth Records.*

4.2.4.3.1.5 Difficulty of Mother's Maiden Name for OV

Across seeker participants, *Mother's Maiden Name* was perceived as quite difficult with a mean average difficulty rating of 2.00 with 1 being *Impossible* and 6 being *Very Easy.* Appendix 139 describes the number of participants selecting each degree of difficulty. Compared to the survey sample, successful participants rated the difficulty as 4.33, perceiving the question as much easier compared to the survey as a whole. Of the 6

successful answers, 4 rated the difficulty as *4 – Neutral* and 2 rated it as *5 – Easy*. Descriptive statistics by accuracy group are provided in Appendix 140.

4.2.4.3.2 Nickname for OV

4.2.4.3.2.1 Accuracy of Nickname for OV

OV's *Nickname* was correctly identified by 13 participants. 40 guesses were recorded. Descriptive statistics of accuracy groups are provided in Appendix 141.

4.2.4.2.2.2 Familiarity with Search of Nickname for OV

For participants who answered correctly, self-assessed familiarity with internet search ranged from (1) *Not Familiar* to (5) *Extremely Familiar,* with a mean average of 2.93. Appendix 142 provides frequency groupings of Likert scale selections. Appendix 143 reports descriptive statistics for familiarity by category. Correct answers are self-critical compared to other groups.

4.2.4.3.2.3 Time of Nickname for OV

Time estimates for searches of *Nickname* varied from 1 minutes to 60 minutes. Appendix 143 describes search time by accuracy grouping. Notably, successful participants reported lower search times than incorrect groups.

4.2.4.3.2.4 Location of Nickname for OV

Nickname for OV was searched for and discovered in a variety of locations. 13 total distinct locations were reported across 43 participants. The remaining participants were unable to located the data. Facebook was, once again, the most popular search location, followed by *Google, Twitter,* and *WhitePages.* Appendix 144 provides a complete list of the search locations across the total survey along with frequency measures for this particular personal data point.

When compared to accuracy, 11 correct answers were located on social media source *Facebook*. Social media in the form of *LinkedIn and Twitter*. Finally, unspecified *"Birth Records"* also provided accurate results.

4.2.4.3.2.5 Difficulty of Nickname for OV

Across seeker participants, *Nickname* was perceived as between *Very Difficult* and *Difficult* with a mean average difficulty rating of 2.67 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 145 describes the number of participants selecting each degree of difficulty. Appendix 146 provides a description of the difficulty rating compared to accuracy groups. Compared to the survey sample, successful participants rated the difficulty as more difficult compared to unsuccessful participants.

4.2.4.3.3 Children's Names for OV

The results of searches for OV's children's names follows. As described in the survey, participants were instructed to report "None" if OV had no children. The results of accuracy, time, location, and difficulty are described and individually compared for accuracy between successful and unsuccessful searches.

4.2.4.3.3.1 Accuracy of Children's Names for OV

Of the participants who supplied answers, the majority were able to correctly determine that OV had no children. A total of 40 participants answered *Children's Names* correctly. The remaining participants reported not being able to find the answer. No participants reported skipping the question. Frequencies of accuracy groups are available in Appendix 147. Descriptive statistics are provided in Appendix 148.

4.2.2.3.3.2 Familiarity with Search of Children's Names for OV

For participants who answered correctly, self-assessed familiarity with internet search ranged from (1) *Not Familiar* to (5) *Extremely Familiar,* with a mean average of 3.05, which is a slightly more confident self-assessment compared unsuccessful groups. Descriptive statistics are provided in Appendix 149.

4.2.4.3.3.3 Time of Children's Names for OV

Time estimates for searches of *Children's Names* varied from 1 minute to 60 minutes. On average, 51 reporting participants provided a mean average of 14.29 minutes spent searching.

When comparing time to accuracy, those participants correctly guessing the *Children's Names* reported search times of 1-60 minutes with an average time of 14.08 minutes, longer than incorrect groups and shorter than not found groups. Descriptive statistics of times compared by accuracy group are provided in Appendix 150.

4.2.4.3.3.4 Location of Children's Names for OV

Children's Name was searched for and discovered primarily on Facebook. 10 total distinct locations for data were reported. *Facebook* was the most popular search location, followed by *White Pages, Twitter,* and *Google. "Lack of evidence"* and "*Search on multiple sites".* Appendix 151 provides a complete list of the search locations across the total survey along with frequency measures for this particular personal data point.

Accurate answers were supplied from social media including *Facebook and Twitter and* commercial venues *Google, Quanki, Spokeo.* Unspecified, *"Lack of evidence",* and *"Search on multiple sites"* were also mentioned by accurate seekers.

4.2.4.3.3.5 Difficulty of Children's Names for OV

Across seeker participants, *Children's Names* was perceived as *Difficult* with a mean average difficulty rating of 3.06 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 152 describes the number of participants selecting each degree of difficulty. Appendix 153 provides descriptive statistics by accuracy grouping. Compared to the survey sample, successful participants rated the difficulty as 3.28 or easier compared to the incorrect accuracy group.

4.2.4.3.4 Pet's Names for OV

4.2.4.3.4.1 Accuracy of Pet's Names for OV

Pet's Names were successfully discovered, second only to *Children's Names.* Like *Children's Names,* seeker participants were instructed to reply "None" if they believed the source participant had no pets in answer to the *Pet's Names* question. A total of 35 participants answered *Pet's Names* correctly. Percentages of correct and accurate answers are provided in Appendix 154. Every guess provided was correct.

4.2.4.3.4.2 Familiarity with Search of Pet's Names for OV

For participants who answered correctly, self-assessed familiarity with internet search ranged from (1) *Not Familiar* to (5) *Extremely Familiar*, with mean average of 3.06, which is a slightly more confident self-assessment compared to other participant groups. Appendix 155 provides descriptive statistics of familiarity.

4.2.4.3.4.3 Time of Search of Pet's Names for OV

Time estimates for searches of *Pet's Names* varied from 2 minutes to 60 minutes with a mean average of 12.86 minutes. When comparing time to accuracy, those participants correctly guessing the *Pet's Names* search times of 2-60 minutes with an

average time of 13.97 minutes. Appendix 156 provides descriptive statistics for time by accuracy group.

4.2.4.3.4.4 Location of Pet's Name for OV

Pet's Name was searched for and discovered in a variety of locations. 8 total distinct locations were reported. Social media venue *Facebook* continued to be the most popular search location, followed by additional social media *Twitter*, nonspecific, "*His Blog"* and *LinkedIn.* The commercial source *Google* was also a frequently reported source. Two participants also added *"no evidence on multiple sites"* in their responses. Appendix 157 provides a complete list of the search locations across the total survey along with frequency measures for this particular personal data point. As all given answers were correct, Appendix 157 reflects locations for correct participants as well.

4.2.4.3.4.5 Difficulty of Pet's Names for OV

Across seeker participants, *Pet's Names* was perceived as quite difficult with a mean average difficulty rating of 2.41 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 158 described the number of participants selecting each degree of difficulty. Descriptive statistics by accuracy group are provided in Appendix 159. Compared to the survey sample, successful participants rated the difficulty 2.86.

4.2.4.3.5 Middle Name for OV

4.2.4.5.5.1 Accuracy of Middle Name for OV

Middle Name was answered correctly by 37 participants. A total of 48 guesses were provided with a total of 14 unique guesses. Percentages of correct and accurate answers are provided in Appendix 160. It is interesting to note that, while incorrect, 13 additional participants answered with the correct initial and, except for 3 seekers, all seeker

participants selected a name which started with the correct initial. No participants reported skipping the question.

4.2.4.5.5.2 Familiarity with Search of Middle Name for OV

For participants who answered correctly, self-assessed familiarity with internet search ranged from (1) *Not Familiar* to (5) *Extremely Familia*r with an average of 2.92, slightly more self-critical compared to individuals who reported not finding the information. Appendix 161 provides descriptive statistics by category.

4.2.4.5.5.3 Time of Search of Middle Name for OV

Time estimates for searches of *Middle Name* varied widely, from 1 minute to 56 minutes. On average participants reported a mean average of 13.65 minutes spent searching. Appendix 162 provides descriptive statistics by accuracy groups.

When comparing time to accuracy, those participants correctly guessing the *Middle Name* reported search times of 2-45 minutes with an average time of 12.08 minutes. Successful search time was shorter than incorrect search time.

4.2.4.5.5.4 Location of Middle Name for OV

Middle Name was searched for and discovered in a variety of locations. 8 total distinct locations for data were reported. *White Pages* is the most popular search location, followed by *Facebook, Twitter Google, LinkedIn* and *University of Washington.* Two participants also stated they "guessed" based on knowledge of the first initial – one of these guesses was correct and one was incorrect. Appendix 163 provides a list of the search locations across the total survey along with frequency measures for this particular personal data point. Correct answers were located from commercial sources *White Pages, Been*

Verified, an employer, *University of Washington.* In addition, *Academia.edu*, and *"Guess"* provided accurate information.

4.2.4.5.5.5 Difficulty of Search of Mother's Maiden Name for OV

Across seeker participants, *Middle Name* was perceived as *"Difficult"* with an average difficulty rating of 3.37 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 164 describes the number of participants selecting each degree of difficulty. Appendix 165 provides descriptive statistics of difficulty ratings by accuracy group. Compared to the survey sample, successful participants rated the difficulty as 3.71.

4.2.4.3.6 Mobile Phone Number for OV

4.2.4.3.6.1 Accuracy of Mobile Phone Number for OV

Mobile Phone Number proved to be a difficult personal data point. A total of 6 participants answered *Mobile Phone Number* correctly. A total of 14 guesses were provided with a total of 10 unique guesses. Percentages of correct and accurate answers are provided in Appendix 166. One of the guesses included the participants physical address. When responding, the source participant noted the information as correct with the annotation *"a little scary".*

4.2.4.3.6.2 Familiarity with Search of Mobile Phone Number for OV

For participants who answered correctly, self-assessed familiarity with internet search ranged from (1) *Not Familiar* to (5) *Extremely Familiar,* with a mean average of 3.00. The more successful participant was slightly self-confident than Not Found Group participants. Appendix 167 describes familiarity in terms of accuracy groups.

4.2.4.3.6.3 Time of Search for Mobile Phone Number for OV

Time estimates for searches of *Mobile Phone Number* varied from 1 minute to 120 minutes. On average, 44 participants reported a mean average of 21.77 minutes spent searching.

When comparing time to accuracy, those participants correctly guessing the *Mobile Phone Number* reported search times of 5-90 minutes with an average time of 25 minutes. Time results by accuracy group are provided in Appendix 168.

4.2.4.3.6.4 Location of Mobile Phone for OV

Mobile Phone Number was searched for and discovered in a variety of locations. 10 total distinct locations for data were reported. *Facebook* remained the most popular search location, followed by *White Pages* and *Google.* Appendix 169 provides a complete list of the search locations across the total survey along with frequency measures for this particular personal data point. When compared to accuracy, correct answers were found in a wide variety of locations including social media venue *Facebook* and commercial resoruces *Pipl, Spokeo, Been Verified, and White Pages.*

4.2.4.3.6.5 Difficulty of Mobile Phone Number for OV

Across seeker participants, *Mobile Phone Number* was perceived as quite difficult with a mean average difficulty rating of 1.65 with 1 being *"Impossible"* and 6 being "*Very Easy".* Appendix 170 describes the number of participants selecting each degree of difficulty. Appendix 171 provides descriptive statistics by accuracy group. Compared to the survey sample, successful participants rated the difficulty as 2.67, perceiving the question as much easier compared to the survey as a whole.

4.2.4.3.7 OV ANOVA

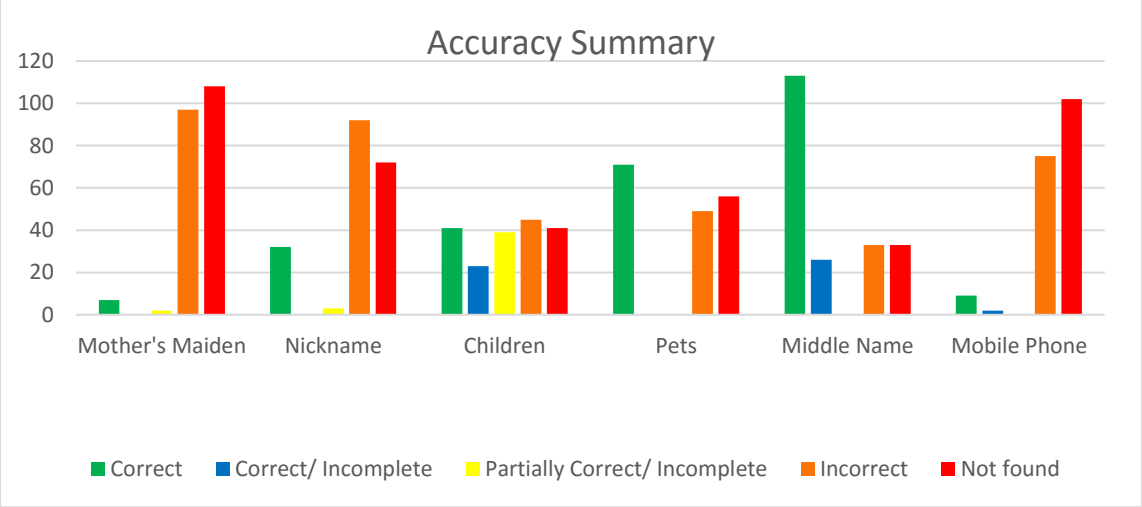
Due to relatively small number of *Correct* groups in several personal data categories, the ANOVA analysis does not consistently represent a statistically significant

finding. Analysis of ANOVAs primarily revealed statistically significant differences in difficulty scores when comparing accuracy groups. ANOVAs for OV may be found in Appendices 172-176.

4.3 Results Summary

The results provide insight into the research questions revealing the relative availability of particular data points described by individual and by data point. The results also describe the difficulty of locating data as well as the location of correctly identified data. The results in relationship to each research question are briefly summarized by personal data point compared to seeker participant accuracy in Table 35.

Table 35. Accuracy of Personal Data Summary



4.3.1 Mother's Maiden Name

Data regarding the *Mother's Maiden Name* was correctly identified for 3 of 4 source participants. In all instances, the number of correct identifications was quite low, ranging from 1-4 participants correctly identifying the *Mother's Maiden Name* of the source participant. This particular question proved difficult as demonstrated in Likert scale ratings and in the amount of time seeker participants reported looking for data as well as in

reference to their overall success. The difficulty rating overall for this data point was 1.98 across the participants. Furthermore, reported search times were longer for this data point than another other data point across all source participants, although for OV, mobile phone number is a very close second.

Locations of correct identification were from government entities and commercial enterprises. Official state record systems, commercial birth record systems and obituary notices were utilized to provide accurate answers. Although social media, particularly Facebook, was the most frequently cited source, in no case did Facebook or other social media sources reveal correct information.

A gender discrepancy was also noted between the participants. Female participants' data were correctly identified only once while that data for the male source participants was identified more often.

4.3.2 Nicknames

Nicknames were identified accurately 3, 4, 10 and 15 times. Both the lower numbers pertain to the female participants. Both the older male (10 correct answers) and the younger male (15 correct answers) were identified correctly more often compared to females. One participant denied having a nickname, which was correctly identified. Seeker participants seemed reluctant to conclude that this source participant did not have a nickname, with only 3 correctly answering *"None"* and the remainder either providing an incorrect answer or reporting "*Not found".* For the remaining three participants, correct information was identified. The difficulty scale was rated at 2.24 across all seeker and source participants. Mean averages across source participants were relatively similar with two participants mean times reported as between 14-15 minutes and the remaining two around 18 minutes.

Locations of correct information identification for *Nicknames* for the female source participants originated exclusively on social media site, *Facebook.* Social media including *Facebook* and *Rate My Professor* provided correct answers for KT along with employer, *Life Pacific College,* and commercial sources *Google,* and *Truthfinder*. While *Facebook* was the most popular search location and also the most frequently cited for correct answers, did not fare as well as other sources when it comes to rate of accuracy. Social media was responsible for correct data locations for OV and included *Facebook, LinkedIn* and *Twitter* in addition to unspecified *"Birth Records".*

4.3.3 Children's Names

Children's names were most often only partially identified. 3 of 4 source participants have children. For the source participant without children, the correct answer, "none" was correctly identified by 40 of 51 seeker participants. The remaining participants, with three to four children each, were all at least partially identified. However, correctly naming all the children without introducing errors and without missing any children was quite rare. For GC and DK, no participant correctly identified all children. 1 participant correctly identified all of KT's children without introducing errors. *Partially Correct/Incomplete* or *Correct/Incomplete* data was identified much more frequently for all three participants with children. Children's names were rated as surprisingly easy compared to accuracy with a mean rating of 2.86. Time on task was often similar with other data points ranging from approximately 12 to 18 minutes on average across source participants.

Children's names were successfully found in a variety of locations. The only correct answer was found via a commercial source, *intelius.com.* For source participant KT, other

categories which contain some level of accuracy were derived from a variety of sources included social media *Facebook and LinkedIn.* Additionally, commercial venues *Truthfinder, Been Verified, White Pages, People search, Obituary, Quanki, LinkedIn, Nuwber, Public Records 360, and* genealogy focused venue *Ancestry* provided accurate records. For source participants GC and DK all answers which were accurate in any measure were derived from social media site *Facebook.*

4.3.4 Pet's Names

Pet's names, for the source participant who reported possessing a pet, were often difficult to find. Only one seeker participant correctly identified the pet name for source participant KT. Participant GC reported previously possessing a pet, whose name was located, however this answer was not current. For participants not owning a pet, particularly DK and OV, seeker participants were often comfortable reporting the *"Correct"* answer, "*None*". It is noteworthy that these two instances did not contain any "*Incorrect"* answers. Seeker participants were either confident enough to report *"None"* or did not find the answer but compared to other data points, they did not report wrong guesses. Compared to other data points, Pet's names were perceived as difficult with a reported difficulty level of 2.23. Time for search ranged from around 12-15 minutes which is on the lower end of the time spectrum considered as a whole.

Social media site *Facebook* was the most frequently reported source of information for accurate answers as well as the only source of a correct answer of an actual pet's names. It was also the most cited as a source for lack of evidence leading participants to conclude that the source participant did not have a pet. Several seeker participants gave insight into their search process as well by reporting locations such as, "lack of evidence on social

media". Other locations reported in connection with accurate answers include social media *blogs, Snap chat, Twitter, unspecified "social media" and twitter.* Commercial sources *Bing, Google, My Heritage, Pipl, Spokeo, Snapchat, Truthfinder and Yahoo* also yielded accurate results.

4.3.5 Middle Name

Middle name represented the data point most consistently and easily found across the source participants. Between 20-35 seeker participants accurately reported the source participants' middle name for three of four source participants. This number also more than doubles the accuracy of any other personal data point for three of four source participants. Middle name was frequently rated the easiest with an average rating of 3.36. Search times ranged from approximately 9 to 16 minutes.

Middle name data was sources accurately from a wide variety of locations including employers, government records, social media and commercial enterprises. It was by far the most accurately discovered data and was accurately discovered in a very wide array of venues.

4.3.6 Mobile Phone Number

Mobile phone number often proved difficult to locate. For two source participants, mobile phone number was not located correctly. In one of those instances, the last four numbers were correctly identified. For the remaining participants, Mobile phone number was located 4-5 times, both slightly more than *Mother's Maiden Name* for the same two participants. In both instances, the *Mobile Phone Number* was successfully located for the male source participants. Difficulty ratings were consistently low for this personal data point at 2.06. Search times ranged from 8-22 minutes on average.

Correct answers were located on social media venue *Youtube* for one participant. In reviewing the link submitted, the video was observed to be an introductory video prepared for a course in which the source participant supplied his mobile phone number. Commercial venues *Been Verified, Facebook, Pipl, Spokeo,* and *WhitePages* each provided a correct answer for the second source participant whose data was successfully located. Finally, a partially correct answer was located using commercial venue *Truthfinder*.

4.3.7 Results Conclusion

Research question one focuses on the ability of a stranger to correctly identify personal information about an individual. Across four source participants, between fifty and sixty participants were able to identify some, but not all, personal data commonly used in authentication. Distinct differences between the data points were noted in the frequency of accurate identification. Significantly, there are several data points that are among the more difficult across all four participants, particularly *Mother's Maiden Name* and *Mobile Phone Number*. *Children's names* were also notably difficulty when it was necessary to correctly identify an entire set without introducing errors.

Research question two questions the relative difficulty of finding the personal data points. Data did show differences in the difficulty of correct identification. This difficulty often corresponds to the accuracy rate of the data point. Frequently, longer search times are observed with data that is not identified correctly. This suggests that perhaps seekers were aware that they were not finding accurate information or perhaps that they were not as adept at the search process. At the easy end of the scale, the Likert scale of difficulty often corresponded with the easiest data point from an accuracy perspective.

Finally, research question three sought to understand the locations of correctly identified data. It was noted that while social media was the most searched location, in

most instances social media did not yield with the most accurate results. Rather, government records and commercial information sources such as *White Pages* often yielded the most accurate results, particularly with the most difficult data points such as *Mother's Maiden Name.*

Chapter 5: Conclusion

5.1 Introduction

The availability, location and difficulty of retrieving personal data are related to personal security in the digital age. As described in Chapter 2, there is a strong connection between personal data and many types of authentication and secondary authentication. Given the difficulty of implementing new authentication systems and the reliance of authentication on personal data, this dissertation examines of the availability, location and difficulty of finding personal data. These findings are of interest to researchers interested in developing new authentication systems and strengthening existing systems. Additionally, industry security personal interested in may leverage the findings to improve protection of their own systems. Government and social agencies interested in establishing norms and guidelines for privacy in public records and the implications of law can benefit from understanding the implications of public records and privacy and security. The findings presented are useful to users concerned with their own privacy as well.

In the following discussions, findings are described with regard to availability, location and difficulty of obtaining personal data to answer the research question. The theoretical and systems design implications are described with respect to each research question and then considered as a whole. Finally, the limitations of the current study are discussed along with recommendations for future research.

5.2 Research Questions

5.2.1 Research Question 1 – Availability

The study found that the majority of personal data sought was successfully located, but not by the majority of seeker participants in most instances. In examining the data, it is

safest to assume generally that information can be found. However, patterns in data accuracy do reveal ways to make personal data more difficult to ascertain accurately.

Some data were accurately located less often than others. The *Mother's Maiden Names* were correctly identified only once for the younger female participant and not at all for the older female participant. *Mother's Maiden Names* were identified for both male participants. Likewise, *Mobile Phone Numbers* were not completely and correctly located for either female source participant but were both located for male source participants. Age of the source participant did not appear to be a factor in this study although age cannot be ruled out as factor. More study is necessary to determine whether gender is specifically a factor in information availability but the results of this study may highlight an interesting area for more research.

Children's Names, particularly in their entirety in cases of multiple children, were rarely fully identified. *Pet's Names* and *Nicknames* were sometimes identified and *Middle Name* was clearly the data point obtained accurately most often. These observations provide insight into the relative obscurity of particular data points. Some data may be more difficult for strangers to find than others and thus may be a safer option when considering personal data used in authentication. It should be noted that these personal data points are still vulnerable to acquaintances and that this study only used strangers to the source participants. Also, no data point examined was entirely impervious to discovery.

For designers and users, it would be prudent to avoid studiously the more available forms of personal data. The scope of the study limits the specific personal data points explores but the results illustrated that there are appreciable differences in the relative availability of personal data which can be leveraged for authentication. This difference

should be considered when selecting personal data for authentication purposes. Additionally, the understanding of the availability of personal data should be regularly updated and augmented based on empirical examinations of web data.

5.2.2 Research Question 2- Location

Research question two explores the locations of found data. Much personal data are self-propagated online in the form of various social media. Popular social media venue *Facebook* was the most searched location in the study. Yet social media and other forums for user supplied data did not yield the most accurate data for the majority of the study. Government resources were rarely used within the context of this study but, when used, often yielded correct results for *Mother's Maiden Name* data point. In no instance was a mother's maiden name accurately identified through the use of social media.

In one instance, a *Mobile Phone Number* for a source participant was correctly obtained from a social media video sharing venue. While self-posted via social media, the video was an introductory course lesson taught by the participant in his employment role. The video included his mobile phone number as a contact point for students. While the professor's availability to students is commendable, the presence of his personal mobile phone number does raise questions as to the support provided by employees for employees. If individual data privacy protections become more important in the future, it will be necessary to consider alternative ways of providing contact and publishing data, such as universities providing phone numbers for adjunct faculty, mobile phone numbers for faculty, or easy to use resources video content management for courses. These types of locations highlight the interplay between actors in making content available.

The remaining source of *Mobile Phone Number* were commercial sites such as phone book resources like *White Pages*. *Children's Names* were likewise only found in completeness using a commercial venues, specifically *intelius.com*. *Middle names* were available from a wide variety of sources including employers, government and commercial enterprises as well as from social media. In some instances, data were discovered on highlighted on social media by employers' use of employee data as part of marketing as well as on publicly available staff and contact pages. When considered in comparison to the number of searches conducted, sources other than social media consistently yielded higher quality results for many types of data points.

The implications of the location data for a substantial portion of the personal data points are significant. The data are often located in venues outside of the source participant's direct, and perhaps even indirect, control. Control, as discussed in Chapter 2, is a significant component to privacy. When considering the ramifications of personal data online and systems security, it is necessary to recognize that single users do not represent a single point of failure that render themselves vulnerable to attack. Instead, it should be recognized that personal data are used in authentication by design of information systems and are made available by information systems outside of the control of the user.

Therefore, it cannot be the responsibility of users alone to protect their own personal data. This study should highlight the need to consider the implications of widely available personal data for authentication and security. Government entities, employers, and commercial venues must deeply consider their information management to determine whether information needs to be available publicly and consider the distribution of that data.

In one instance in the study, data were revealed in a form letter sent to parents in an elementary public-school system. This letter was readily found online and fully available. This availability is clearly outside the scope of "need to know" and, at one time, would have been sent in physical form to parents and students. Now, it is readily available online, perhaps in perpetuity. Furthermore, it reveals information which accurately identified personal data for one of the source participants. This type of information availability illustrates the types of considerations for information distribution which need to be more deeply considered by designers, businesses, government entities. Limitations and ownership of commercial data should also be considered from the perspective of law. This study informs that decision making by revealing a significant perpetration of personal data availability which is beyond the scope and control of the individual.

The personal data which were more readily located on social media and other forms of self-propagated media include partial data for *Children's Names, Pet's Names* and *Nicknames.* It is encouraging that only partial information was found in several instances and that *Pet's Names* were not found readily, although this study may be impacted by the lack of pets possessed by source participants. *Nicknames* were widely available from both self-propagated social media and commercial venues.

Users should consider their own password and secondary authentication practices. The study was limited by the pet ownership of the source participants. Only one source participant currently owns a pet. *Pet's Name* was correctly located using social media venue *Facebook* by one seeker participant. *Facebook* was also a source of outdated information. In one instance, a participant noted that she had previously owned a pet, whose name was found on *Facebook*. Considering that information contained in passwords and

secondary authentication is, sometimes outdated, the availability of historical information is also interesting. This result suggests the need to examine the use of historical data for authentication. Given the availability of web data, historical data used in authentication may not be rendered obscure as intended.

Individual users should also be made aware that certain data points were more likely to be found via social media are may, therefore, be more likely to be under their control to some degree. Data such *as Middle names, Nicknames, Pet's Names and Children's Names* were often obtained, at least in part, via social media venues. Individual awareness of personal data availability which may remain in the control of the user becomes important to security conscious users wishing to create secure passwords and in making recommendations to users regarding password content for memorability and obscurity.

5.2.3 Research Question 3 - Difficulty

From a difficulty perspective, mother's maiden name and mobile phone number are clearly the more difficult data points when considering the entire survey. Finding complete information regarding multiple children also proved difficult, although partial and incomplete information was often reported correctly. *Middle Name* was also clearly the data point obtained accurately most often. These extremes provide insight into what may be safer, although not entirely secure. These data points were more obscure in this study and thus may be safer options than middle name or nicknames. *Nicknames* and *Pet's Names* were not as difficult to identify as *Mobile Phone Number* or *Mother's Maiden Name* but were more difficult than *Middle Names.*

In instances where information consisted of multiple correct answers, such as children's names, complete information was rarely found. It is also noted that while many

participants accurately identified several data points, in no instances did any seeker participant accurately identify all six data points. These points may suggest that information security may be increased by combining multiple forms of the personal data which is more difficult to obtain, greatly increasing the difficulty of accurate, complete information availability. Of course, this security introduces opportunities for user error as well as promoting poor interaction design from a usability perspective. Unfortunately, the number of data points required to improve security, based on this study, is likely be higher than three or four, which becomes untenable from a usability perspective. The requirement for ease of use and the limitations to cognitive load generally necessitate a limited interactions for task completion as well as a limited amount of recalled knowledge (Besnard & Arief, 2004). The concept of a threshold of difficulty for a particular number of personal data points for authentication deserves consideration in future studies.

For designers and users, it would be prudent to avoid studiously very available forms of personal data. The scope of the study limits the specific personal data points but the results illustrate appreciable differences in the relative accessibility of personal data which can be leveraged for authentication. These differences should be considered when selecting personal data for authentication purposes. The relative availability of personal data also deserves ongoing study as the web continues to develop. Designers and users may be able to increase protection by requiring more complete forms of data that contain multiple parts. However, usability concerns will be a key factor in the practicality of this approach based on cognitive limitations.

5.3 Summary of Findings

Generally, the results demonstrate the fundamental flaws of designing security systems constructed around personal data in the digital age. The internet and web exist to make information readily available. In the interest of digitizing, the benefits of information obscurity were overlooked. In this case, information obscurity aids in security due to the design of secondary authentication systems and the cognitive limitations of password based authentication. Security based on personal information is widely entrenched, as discussed in Chapter 2, and as a result systems are compromised by inattention to the changing availability of data. While this study does point to some hopeful mitigating factors, in general, the dissertation demonstrates the fundamental flaw of designs based on personal information.

When considering Actor Network Theory as it applies to security (Corbett, 2013; Hartzog & Stutzman, 2013; Powell, 2011), the data suggest that individual users may be less culpable in the propagation of their own personal data than is often imagined. The data show that the majority of correctly identified data did not originate directly from user propagation but the data were firmly in the hands of secondary or tertiary actor. The data were in the possession of and propagated by companies, employers or government organizations. Even in instances where the data were self-propagated, such as the mobile phone number on Youtube, the information was provided as an aspect of employment and connected to work activities. (Belanger & Crossier, 2011) suggest that control is a fundamental aspect of privacy. In this study, the data suggest the lack of control on the part of the source participants. While fundamental responsibility for security is often laid upon the "weakest link", these data suggest that a significantly overlooked weak link is the

propagation of personal data clearly outside the direct control of the individual. As such, the responsibility for reasonably discriminating data and providing adequate data privacy falls to the propagating entities. For a meaningful discussion of personal privacy to continue, it is necessary to reengage individuals in direct control of their information or evaluate the concept of privacy based on control.

It is worth noting that a good deal of the information explored would not necessarily be considered private by some. For example, records of marriage, and hence, the ability to track maiden names predates digital data collection. Likewise, information such as obituaries, recorded by local newspapers, while recorded, were not globally available until the advent of modern data infrastructure. Thus, the data would have required an arduous amount of research and expenditure to obtain. While the vital information was considered public, it was also rendered obscure and difficult to collect *en masse* due to the nature of pre-digital and pre-web record storage and information system processes.

Obscurity is considered a fundamental issue in modern privacy discussions (Hartzog & Stutzman, 2013). The obscurity of records, locally recorded and often only available in quantities that could be hand recorded only in a particular location have been replaced by searchable government databases. The enigmatic nature of personal data was preserved by factors such as archaic records systems, geographical considerations, and volume. Now, systems are easier to search and access and are thus rendered potentially susceptible to attacks based the ready availability of personal data.

While it would be inappropriate to read more into the findings than are indicated, the data suggest that a new understanding of data privacy is necessary. The continued understanding of individual privacy cannot rest on concepts of individual control. as the

data are not in the control of the individual. Active social and global forces are necessary to restore individual privacy to the control of the individual. The practical complications of attempting to restore power to the individual cannot be overestimated in light of factors such as the great economic benefit possessing data, the global nature of information systems and the difficulty of retracting data from current systems.

Obscurity may offer a modicum of hope for privacy, as demonstrated in this study. Certain data are more difficult to obtain than others. However, as nearly all data were obtained, obscurity does not provide meaningful privacy protection without additional measures. Intentionally obscuring data runs counter-intuitive to the ethos of the digital web-data based mentality which prizes decentralized autonomy and the right to post any information. The right to post any desired information and the extent to which this right extends to organizations and individuals' perpetuation of data requires scrutiny particularly as it pertains to data regarding or belonging to other individuals. Furthermore, in current web ideation, organization and individuals are permitted to collect data nearly indiscriminately. This intimates that privacy in respect to individual control and obscurity cannot be maintained without introducing some form of centralized authority. The fundamental incongruous result of freedom of information on the web is the subjugation of individual control of personal data.

In the ongoing debate regarding the very meaning of privacy and the implications of privacy concepts in information systems, it necessary to consider the implications of personal freedom on personal data collection, storage, dissemination and propagation. More research is needed to arrive at a meaningful understanding of privacy in the digital age. The implications of this dissertation on privacy and security using personal data are

profound, but by no means conclusive or exhaustive. The findings suggest challenges to current concepts of privacy discussed in Chapter 1 and evoke a recondite discussion of the meaning of privacy which is devoid of control and lacking obscurity.

The findings also suggest that it may be necessary to determine what data, if any, should be considered private and whether any data deserve an exemption from the Web principles of freely posting data without constraint. The form of this constraint and restraint is also a necessary discussion. Centralized constraint would hinder the purpose of the web and arguably constrain individual freedom while personal restraint cannot protect privacy of individuals who do control personal data regarding themselves. The nature of personal data on the web and the relationship between freedom of expression and personal privacy remain at the crux of the discussion of privacy on the Web. In light of the findings of this study, the IS community needs to deeply consider the original goals of the Web. More research is required to meaningfully discuss the psychological impacts of privacy (Jourard, 1966), the debate around the legal of privacy (Borchert et al., 2014; Bunn, 2015; Hartzog & Stutzman, 2013; Palmer, 2011; Powell, 2011) and its implication on the continued development of information systems.

5.4 Limitations

The study was limited in several regards. IRB protection for source participant was considered a very high priority for this study and so the design was created with the idea of protecting any data which are not already available. With this in mind, personal data were not collected from the source participant in advance of the study so that any personal data which were not discovered would remain private. As a result of this study design, it was not feasible to reward or otherwise motivate seeker participants when they provided a

correct answer. In a real-world scenario, personal data collections with regard to authentication would presumably yield some form of reward for accurate information. This factor was not replicated in in this study. In some instances, seeker participants were rewarded in the form of IRB approved extra credit subject to instructor approval. The majority of seeker participants were undergraduate students completing the survey for extra credit in a course. Seeker participants were not rewarded for correct answers which may impact their motivation to provide correct answers.

Like seeker participants, source participants were all volunteers. The privacy perspectives, attitudes and behaviors of the source participants were not explored. The impact, if any, of privacy attitudes of the source participants on data availability was not included in the research or study design and is beyond the scope of the research questions. Based upon the sources of information used to inform accurate answers, it is questionable whether privacy attitudes would impact information availability.

As discussed in the literature review in Chapter 2, previous studies aimed at measuring knowledge of personal information used lab settings. Schechter, Brush, and Egelman (2009) found that the lab setting may limit the amount of time participants are willing to look for information. In this study, seeker participants were permitted to work through the study at their own pace and from a location of their choosing. This design choice was made to mitigate limitations of previous research but also introduced the limitation that the searches themselves were not observable. As a result, for purposes of the study, time frames should be considered user reported estimates and not timed experiments. The start to finish timed survey results do provide some measure of accuracy check on the amount of time estimated. However, cognitive sciences show that many

factors can impact time perceptions and therefore reported times are limited by seeker participant perceptions. The start to finish timed survey results do provide some measure of mitigation on the amount of time estimated as described in Chapter 4.

5.5 Future Directions

This dissertation provides an empirical understanding of the discoverability of personal data used in authentication. In particular, three aspects were explored: the availability of personal data, the location of discovered data and the relative difficulty of discovery measured by time and perception of difficulty. Understanding the actual, rather than presumed, availability of personal data used in authentication represents a significant contribution to the advancement of information systems. The dissertation provides a meaningful foundation for continued research in developing new systems for authentication and mitigating risks with current systems. The findings contribute to the formation of the foundation for an ontology of personal data available online as it relates to authentication. The locations of personal data provide insight into security risks and also provides potential insight into privacy concepts and the implications on personal privacy of individuals.

Future research seeks to mitigate the limitations of this study by examining accuracy of search under various circumstances, such as reward connected to correct answers. Additional venues for recruitment, such as white hat hacker communities, provides interesting opportunities for exploring the impact of seeker participant expertise on results. A study including explorations of the privacy attitudes of seeker participants to discover whether information sharing behaviors of individuals impact the discoverability

of personal data would greatly contribute to the continued development of understanding privacy and personal information.

Specific interesting trends suggested by the data deserve additional consideration. For example, directly examining the possible connection between gender and the relative obscurity of mother's maiden name. Additionally, the number of accurate answers provided should be explored with additional data points and under other conditions to establish a threshold for data difficulty as related to the volume of personal data required to improve security. Specific methodologies which introduce limitations deserve consideration for future study. Finally, newly proposed authentication techniques and associated personal information may be examined for susceptibility to human search. Additional research is required for confirming and expanding the findings and observing for additional factors which may influence data availability under various circumstances. Additional research which addresses the limitations of this study may be used to meliorate the difficulties associated with personal data used for authentication.

In conclusion, an applicable understanding of six personal data points as they pertain to four individuals is deeply explored in this dissertation study, enabling the formation of a foundational understanding of personal data availability, location and difficulty of discovery by strangers. The implications of the study include both direct and actionable applications to industry as well as informing a broader theory of information systems as it pertains to personal data availability on the web, security as related to personal data used in authentication and foundational concepts of privacy.

Appendices

Appendix 1. Personal Data Survey
Seeker Participant Data Collection

Participant Instructions:

Thank you for your help. I am attempting to understand how easy it is for people to find out data about a stranger using the internet. To complete the survey, please do your best to correctly identify the information requested about [Insert Name] Please feel free to use online search engines, such as Google, Yahoo, Bing or others, social networks, such as Facebook and any other tool you might use to learn information about a friend, family member, favorite sports player or other celebrity or any other individual of interest to you. Do refrain from illegal activity and paid services to acquire this information. Remember that it is illegal to attempt to access an account that belongs to another individual.

If you believe you have found or could correctly guess the information, write your answer to the question. Then provide information about the time you spend looking for that particular piece of information, the location you found the information, and how hard it was to find it. Please see the sample sheet for an example of how to fill out the survey.

If you did not try to find the information, please circle "NOT ATTEMPTED" on the form. For any attempted question, please answer as many of the questions as possible, such as where you looked for the data and how long you looked, even if you are not able to find the information.

This information will help you identify data about [Insert Name]. If you are already acquainted with this individual, please indicate your prior knowledge and skip to the next person.

Name
City, State
Photograph

I am already acquainted with this person. Yes: _____ (skip to next person)
No: _____ (continue on following page)

Question 1
Please briefly describe your familiarity with conducting online searches for information:

1. Not at all familiar with conducting online searches for information
2. Slightly familiar with conducting online searches for information
3. Somewhat familiar with conducting online searches for information
4. Moderately familiar with conducting online searches for information
5. Extremely familiar with conducting online searches for information

Question 2

If you decide to skip this question without trying to find the answer, please circle here:

NOT ATTEMPTED

If you tried to find the information, but did not locate it, please tell us how long you looked (question 3).

1. Sarah Smith's mother's maiden name:

---

2. Location:

---

3. Time (in minutes):

---

4. How difficult was it to find this information? Circle the answer.

    1. Impossible – I looked a lot, but I didn't find it
    2. Very difficult – Almost didn't find it
    3. Difficult – Looked a lot
    4. Neutral – Wasn't particularly hard or easy
    5. Easy – Checked a few places and found it easily
    6. Very easy – Took almost no looking

Question 3

If you decide to skip this question without trying to find the answer, please circle here:

NOT ATTEMPTED

If you tried to find the information, but did not locate it, please tell us how long you looked (question 3).

1. Sarah Smith's Nickname:

---

2. Location:

---

3. Time:

---

4. How difficult was it to find this information? Circle the answer.

1. Impossible – I looked a lot, but I didn't find it
2. Very difficult – Almost didn't find it
3. Difficult – Looked a lot
4. Neutral – Wasn't particularly hard or easy
5. Easy – Checked a few places and found it easily
6. Very easy – Took almost no looking

Question 4

If you decide to skip this question without trying to find the answer, please circle here:

NOT ATTEMPTED

If you tried to find the information, but did not locate it, please tell us how long you looked (question 3).

1. Sarah Smith's child(ren's) names:

---

2. Location:

---

3. Time:

---

4. How difficult was it to find this information? Circle the answer.

1. Impossible – I looked a lot, but I didn't find it
2. Very difficult – Almost didn't find it
3. Difficult – Looked a lot
4. Neutral – Wasn't particularly hard or easy
5. Easy – Checked a few places and found it easily
6. Very easy – Took almost no looking

Question 5

If you decide to skip this question without trying to find the answer, please circle here:

NOT ATTEMPTED

If you tried to find the information, but did not locate it, please tell us how long you looked (question 3).

1. Sarah Smith's Pet(s) names

---

2. Location:

---

3. Time:

---

4. How difficult was it to find this information? Circle the answer.

1. Impossible – I looked a lot, but I didn't find it
2. Very difficult – Almost didn't find it
3. Difficult – Looked a lot
4. Neutral – Wasn't particularly hard or easy
5. Easy – Checked a few places and found it easily
6. Very easy – Took almost no looking

Question 6

If you decide to skip this question without trying to find the answer, please circle here:

NOT ATTEMPTED

If you tried to find the information, but did not locate it, please tell us how long you looked (question 3).

1. Sarah Smith's middle name

---

2. Location:

---

3. Time:

---

4. How difficult was it to find this information? Circle the answer.

1. Impossible – I looked a lot, but I didn't find it
2. Very difficult – Almost didn't find it
3. Difficult – Looked a lot
4. Neutral – Wasn't particularly hard or easy
5. Easy – Checked a few places and found it easily
6. Very easy – Took almost no looking

Question 7

If you decide to skip this question without trying to find the answer, please circle here:

NOT ATTEMPTED

If you tried to find the information, but did not locate it, please tell us how long you looked (question 3).

1. Sarah Smith's mobile phone number:

---

2. Location:

---

3. Time:

---

4. How difficult was it to find this information? Circle the answer.

1. Impossible – I looked a lot, but I didn't find it
2. Very difficult – Almost didn't find it
3. Difficult – Looked a lot
4. Neutral – Wasn't particularly hard or easy
5. Easy – Checked a few places and found it easily
6. Very easy – Took almost no looking

Appendix 2. Source Participant Data Collection

Personal Data Verification Questionnaire

Instructions: for each piece of information, please indicate if the answer provided is correct.

Your Mother's Maiden Name	
Data	Accuracy
Surname	Correct/Incorrect
Surname	Correct/Incorrect

Your Nickname	
Data	Accuracy
Sample Nickname	Correct/Incorrect

Your Children's Names	
Data	Accuracy
Sample Name	Correct/Incorrect
Sample Name	Correct/Incorrect
Sample Name	Correct/Incorrect
No Children	Correct/Incorrect

Your Pet's Names	
Data	Accuracy
Sample Name	
No pets	Correct/Incorrect

Your Middle Name	
Data	Accuracy
Name	Correct/Incorrect

Your Mobile Phone Number	
Data	Accuracy
555-555-5555	Correct/Incorrect

Source Participant Demographics

Your first name: _____
Your last name: _____
Location (City, State): _____
Gender: _____
Age: _____
Upload a photograph here

Appendix 3. IRB Application

Protocol Title: Risk Analysis of the Discoverability of Personal Data Survey

Is this application associated with a Planning and Development activity? If yes, please provide the date the ORPC provided administrative approval, the IRB approval number and title: N/A

If applicable, provide the funding agency, the sponsored project title and UMBC award ID: N/A

List the Principal Investigator(s) below. Please list other research team personnel on Page 3 of this application**.** Students may be listed as an Investigator; faculty advisors must also be shown and sign this form. ***Attach an abridged vita or resume*** to this application highlighting expertise of the Principal Investigator(s) as it relates to this study.

Name	Department	Phone Number	E-mail	Date CITI Education Program was completed *
Kirsten Richards	**Information Systems**	**443-510-8241**	**kirsten5@umbc.edu**	9/15/2014
Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter a date.
Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter a date.

*** If you need information about training completion dates, please contact the ORPC**

Does the Principal Investigator(s) or any of the project personnel have a financial interest related to the research or sponsor (e.g. payment for services, equity interests, etc.) that must be disclosed according to UMBC Conflict of Interest policy?
Yes ☐ No ☒

Type of Review Requested:

Expedited ☒- check the appropriate box at the end of this application
Full Board ☐ - complete the necessary information on Page 7 of this application

Electronically submit the protocol and any accompanying documents to irbsubmissions@umbc.edu.

By typing your name, email address and date, the investigator(s) certify they will abide by all UMBC IRB policies and procedures and understand that no research activities will be conducted with human participants prior to obtaining the required approvals. The investigator(s) will inform the IRB at the earliest possible date of (1) any significant changes in the project with respect to human subject participation, (Kaikkonen et al.) any adverse reactions or unexpected responses observed involving human participants, and (3) any need for continuation of the project activities beyond the approval date. Faculty advisors who type their name, email address and date certify they have read and reviewed this proposal and confirm it is ready for review by the IRB. Faculty advisors agree to mentor the student during the term of IRB approval.

Investigator's Signature: Kirsten Richards Email: kirsten5@umbc.edu Date: 5/16/2016

Investigator's Signature: Click here to enter text. Email: Click here to enter text. Date: Click here to enter a date.

Faculty Advisor's Signature: Click here to enter text. Email: Click here to enter text. Date: Click here to enter a date.

IRB Action: ***Expedited _____ Full Board Review _____***

Approved - IRB Chair _____ Date _____

(application for approval of use human participants form) –08/19/2015

1) <u>Anticipated start date of the research:</u> 6/13/2016

<u>Approximately how long will it take to complete the research objectives (months/years): 12-18 months</u>

2) <u>List all other personnel who are working on this study</u>

Name	Role	Department	Phone Number	E-mail	Date CITI Education Program was completed
Anthony F. Norcio	**FA**	**Information Systems**	**410-455-3206**	**norcio@umbc.edu**	6/12/2014
Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter a date.
Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter a date.
Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter a date.
Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter a date.
Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter a date.
Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter a date.
Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter text.	Click here to enter a date.

Role: Faculty Advisor (FA) , Research Assistant (RA), Graduate Student (GS), Undergraduate Student(US)

Will the procedures in this application be used for thesis, masters or dissertation research? Yes ☒ No ☐

If yes, please list thesis or dissertation committee member names: Dr. Anthony Norcio (chair), Dr. Wayne Lutters, Dr. Lina Zhou, Dr. Aaron Massey, Dr. Richard Forno

Planned graduation date? **5/24/2017**

3) <u>Purpose of the Study:</u> What are the specific scientific objectives (aims) of the research? Please attach additional information to this application (i.e. specific aims, project description,etc.) if you wish to provide additional information about the protocol.

This research seeks to address a current knowledge gap in information systems by exploring the discoverability of personal data online. Specifically, the research examine selected personal data points regarding the participants and seek to determine whether those data points are discoverable by other individuals. Three areas will be explored: 1. Whether personal data can be accurately identified using public, online resources, 2. How difficult personal data is to identify, and 3. The web locations identified by participants as sources for access to personal data.

4) Procedures: Describe the all procedures of the study in which human participants will participate. **Please include Microsoft Word versions of recruitment fliers. Adobe Acrobat (.pdf) versions of questionnaires, surveys or other measures related to the proposed project are acceptable**. When using multiple questionnaires, surveys or other measures, describe which questionnaires, surveys or other measures will be used for specific procedures.

The experiment will be conducted in two parts, a pilot study and dissertation study. During the pilot study, the use of a web survey will be compared to the use of a hard copy of the survey. The survey instrument will be otherwise identical and is included as "Personal Data Survey". Additionally, the amount of information necessary to provide participants will be determined. Current research suggests that a name, location and photograph are sufficient to successfully identify individuals online. This belief will be confirmed for use in the subsequent dissertation study. Survey participants will be provided with either name and city/state or name, city/state and photograph of the source participant. The least intrusive method that is sufficient for the study will be used in the subsequent dissertation study.

During the pilot study, two groups of participants will be recruited. Informed consent will be obtained from source participants and then seeker participants. Source participants will be asked to supply their demographic information at the beginning of the study. This form is provided in the attachment as "Source Participant Demographics". Seeker participants will be asked to locate specific personal data points regarding the source participants online. Five points of personal data will be examined including: mother's maiden name, nickname, pet's names, children's names and middle name. These personal data points will pertain to the second group of participants, the source participants. Seeker participants will be asked to identify the data requested, record the amount of time spent searching for data, specify the location of the data on the web, and record their impressions of the difficulty of locating the data. Source participants will subsequently be asked to verify the accuracy of data collected by seeker participants. This data is collected using the "Personal Data Verification Questionnaire".

The dissertation study will follow the same model described for the pilot study. The source participants will provide consent to look for specific personal data points online. They will be informed of the nature of the data collected as well as specific data points. The seeker participants will attempt to correctly identify the

data requested using publicly available web resources. This is collected on the "Personal Data Survey". The "Personal Data Verification Questionnaire" will be used with the source participants to confirm the accuracy of data collected in the survey. The questionnaire will be constructed after the collection of data by seeker participants, using the data provided. A sample of the anticipated construction is provided. No new personal data will be collected from the source questionnaire - that is, any personal data which is not discovered online by participants will not be revealed or collected as a result of the study.

5) <u>Participant selection:</u> Who will be the participants? How and from where will they be obtained? What are the criteria for inclusion and exclusion? What is the estimated number of participants and age range? How will eligibility be determined, and by whom? Will the participants be selected for any specific characteristics, e.g., age, sex, race, ethnic origin, religion, or any social or economic qualifications?

The two groups of participants, source participants and seeker participants, will be obtained separately.

Source participants are purposefully sampled to represent a variety of ages and gender demographics. Approximately four participants will be recruited for the dissertation study from the following age groups: 18-29, 30-49, 50-64 and over 65. These participants will be equally split between male and female participants. This model reflects designs of other studies looking at personal data. One additional participant will be selected for the pilot study. Eligibility will be determined by the primary researcher.

Seeker participants will be recruited openly with posters and will include recruitment of UMBC students. Participants will be screened as part of the survey to determine whether or not they are personally acquainted with source participants. Participants will be shown the names of source participants and asked whether they are acquainted with the particular source participant. Acquainted is defined by the Merriam-Webster as, "someone who is known but who is not a close friend: the state of knowing someone in a personal or social way : the state of knowing someone as an acquaintance" . If participants indicate a knowledge of the source participant by answering "yes" when asked on the survey whether they are acquainted with the source participant, that seeker participant will be excluded from answering questions regarding the participant. The purpose of exclusion is to minimize the influence of prior knowledge obtained outside of the parameters of the study and to better answer the research questions which address the knowledge obtainable from online data sources. The location question on the survey will also assist in discovering whether other knowledge are used to answer the questions.

6) <u>Process of Consent:</u> How and where will the consent process take place? Who, among the research team members, will obtain consent? What information will be provided to participants if a research study deals with anonymous

research, recording instruments or reportable activities (e.g. illegal drug use, child abuse, etc.) What steps will be taken to avoid coercion or undue influence? Describe the process here and make sure the process is consistent with description in the consent or assent forms. If not obtaining written consent (with submission of a waiver of written consent request), how will consent conversations be documented (consent log, spreadsheet, etc.)? Please include with the application.

Please include Microsoft Word versions of all consent and assent documents or consent scripts.

Consent will be obtained from source participants at the beginning of study. A model consent form is included for source participants. Consent will be obtained by the primary researcher. The study will be described to the participants and they be informed of each of the data points to be collected. The data collected will not pertain to reportable activities such as drug use. The data collected will only be data available on the web which may be found by other individuals. Individuals will be informed of the nature of the data and be asked to ensure that none of their passwords reflect the data collected and also be asked to avoid changing their online behavior. The consent for this study will be obtained by the researcher.

Consent will be obtained from the seeker participants as they are recruited. They will be cautioned not to use the data they find beyond the scope of the survey and avoid using illegal and paid services for obtaining data. There is no anticipated potential harm to seeker participants beyond fatigue upon taking the survey. As the survey will be taken entirely upon the seeker participant's willing participation and at their own desired pace, participants will be able to complete the study and rest as needed.

Which consent documents are attached to this application?:

☒Adult Consent Form ☐ Child Assent Form ☐Waiver of Written Consent
☐Oral Consent Script ☐ Telephone Consent Script ☐ Information Sheet ☐ Email Consent Document
☐ Parent/Guardian Consent Form ☐ Web-based Consent Form

7) <u>Data Collection, Storage and Confidentiality:</u> How will data be collected and recorded? Will it be associated with personal identifiers or coded to protect personal privacy? Who will have access to the data and/or to the codes? If data with participant identifiers, who will have or maintain access to this information? If providing payments to participants how will these payments be tracked and identifying information kept secure? If a participant decides to withdraw from this study, what procedures will you use to protect the confidentiality of the data during your analysis? Provide a location where data records or information will be stored or available. Where will data and associated protocol files reside upon completion of the study? Will be use a computer, laptop, tablet or smartphone to collect data?

Data is collected via survey or questionnaire as described. Written survey documents will be deidentified and stored anonymously. Web survey data will be password protected. Each survey will include identifiable information regarding the source participant and may be stored in an encrypted format. Again, only information which is publicly available online will be collected. The data regarding source participants is not coded, as the object of the experiment is to ascertain whether web data can be accurately collected regarding individuals, however, results of the data collected and analyzed as a result of the research will only be reported using deidentified information including statistical descriptions of results and pseudonyms. The final data of the survey will only be directly accessible to the primary researchers. Furthermore, no information will be collected from source participants, aside from name, location and photograph, that is not available publically online. Personal information from seeker participants will include only their self-identified computer experience.

If participants are provided with remuneration for their participation in the study, this information will be tracked using a password protected Excel spreadsheet. Remuneration is not planned at this time, but may become necessary as reflected in other similar studies. Remuneration may take the form of monetary payment or extra credit in classes, as approved by IRB. In the event that extra credit is provided, the names of participants only will be supplied to the appropriate instructor with the permission of the student.

Confidentiality of collection of sensitive information may require investigators to follow appropriate security protocols according to UMBC's Data Use Guidelines. Review the levels of security that may require data protcections. The UMBC Department of Information Technology (DoIT) may be brought in to prepare a risk assessment documents and/or perform an onsite inspection of the PIs' data access and storage facilities. You may be required to provide information on encryption techniques, data access, etc. If so, it ***must be*** detailed in the IRB protocol application for review.

Because some seeker participants may include UMBC students, the names of UMBC students may be collected. However, personal data regarding these individuals, except their level of personal computer experience, is not collected. Furthermore, the results of the survey will be deidentified as regards the seeker participants. Source participants are currently planned for recruitment outside the UMBC community. Participants will be instructed to collect data regarding the source participants only from legal, free web sources. Therefore, the data in the survey and questionnaire is Level 0, as it is obtained from online, public sources.

8) Research that use data, records or human biological specimens *with* direct participant contact (complete if applicable):

What procedures will you and the research team put into place to minimize or eliminate exposure to potentially infectious agents that may be present in the

specimens (i.e. human blood, tissues or body fluids)? Describe your plan for exposure control and personnel protection.

NA

Will the activities of this research fall under the HIPAA Privacy Rule? Yes ☐ No ☒
If "Yes," describe the procedures you will use to comply with the HIPAA Privacy Rule Click here to enter text.

9) <u>Research that use data, records or human biological specimens *without* direct participant contact (complete if applicable):</u> NA

What are the types of data or specimens?

☐ Data already collected for another research study
☐ Medical records
☐ Patient specimens (tissues, blood, serum, surgical discards, etc.)
☐ Other (specify): Click here to enter text.

What is the source of the data or specimens and how were they collected? Describe the process of data collection including consent, if applicable.

Are the data or specimens publicly available? (That is, can the general public obtain the data or specimens? Data are not considered publicly available if access is limited to researchers.)
NA

If the data or specimens are not publicly available, are you required to obtain permission to access these?
Yes ☐ No ☐
If the answer is "yes," attach a copy of the correspondence/or sample data use agreement granting you permission to access and use data.

Will you be receiving data or specimens in an identifiable format or that will remain identifiable in the research records? Yes ☐ No ☐
What confidentiality measures will you put into place to protect identities? Click here to enter text.

Data holders whose archives are available on a restricted basis have certain conditions for use and possession. Investigators (the "data users") must be aware of these provisions as their research must conform with confidentiality and data protection provisions of the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA), the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule and/or the Family Educational Rights and Privacy Act (FERPA). Each of these regulations obligates "data users" to protect the privacy and confidentiality of personal identifiable information that they possess and to obtain permission, when warranted, from individuals to disclose information. Users may also be audited by federal agencies to make sure they are following proper procedures. Penalties for non-compliance with these regulations include financial fines are/or imprisonment.

What procedures will you and the research team put into place to minimize or eliminate exposure to potentially infectious agents that may be present in the specimens (i.e. human blood, tissues or body fluids)?
Describe your plan for exposure control and personnel protection. NA

10) Risks/Benefits: What potential benefits may participants receive as a result of their participation in the research? What are the potential risks/discomforts associated with each intervention or research procedure? What procedure(s) will be utilized to prevent/minimize any potential risks or discomfort?

Source participant's online, public information will be examined in the course of the study. Risks could include illegal or unethical use of this data by seeker participants. The data collected is limited to the specific data in the study as described. Seeker participants will be advised not to use data for any other purpose than the survey. Source participants will also be encouraged to ensure that their personal accounts are protected with strong passwords and will be advised on the creation of strong passwords and secondary authentication systems. This advice is included in the consent. Source participants are informed that their personal data may be discovered by other research study participants. Benefits to source participants will include the outcome of the survey. Participants will be provided with information regarding the data discovered pertaining to themselves online. This will provide the participants with a better understanding of their personal privacy online. Fiscal remuneration is not planned at this time, but may become necessary as reflected in other, similar studies.

Seeker participants are anticipated to experience no potential risks or discomforts except for fatigue upon completing the survey (either web or hard copy based). As the survey is conducted entirely off site and time is not limited, participants may rest as required. Participants will be advised not to use illegal means to attempt to obtain data and particularly advised that attempting to access another individual's online accounts is illegal. Furthermore, they will agree during consent not to use information discovered outside of context of the research study. Seeker participants who are members of the UMBC student body may be offered extra credit in courses for participating, subject to instructor and syllabus rules and approval. If such a benefit is offered, instructors will only be informed of a participants' involvement in the survey.

11) Location: Where will the study be conducted (e.g. institutions, organizations, facilities such schools, churches, child centers, businesses, nursing homes, conferences, etc.). Is local or institutional IRB approval from the recruitment/research site required? If so, please include a copy with the application. Letters of cooperation from sites that generally consist of a broad statement indicating that the researcher will be allowed to recruiting participants, conduct his or her study procedures and collecting data at a specific facility are not considered human subjects use approval but may

be submitted as part of the application.

The participants will participant from self-selected locations which may include home or other venues selected by the participant. Seeker participants will need to use computers to access a web based survey, which is anticipated to be provided via Survey Monkey or a similar service. Physical survies will be mailed or emailed as attachements, based on the preferences of the survey taker, for the pilot study. Source participants will be provided with a list of results, which may be reviewed at their disgression in a location of their choosing. A specific physical site is not required by the survey design.

COMPLETE ONLY FOR MORE THAN MINIMAL RISKS STUDIES

A) Background: Please provide an evaluative summary of relevant literature on the topic **if** your protocol falls within the "**More than Minimal Risk** "category:(defined as: *"where the probability and magnitude of harm or discomfort anticipated in the proposed research are greater, in and of themselves, than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or test". [45 CFR 46.102(i)]*,)"

NA

a.. If adverse effects occurred, indicate how your research is addressing or attempting to prevent such effects. Include full citations for included research. If possible, also include a copy of relevant articles.

NA

b. For **More than Minimal Risk** studies that ALSO include invasive procedures, indicate which databases have been consulted (e.g., Medline). Summarize findings, including findings of adverse effects and steps taken by you to prevent this from occurring in your protocol. You may reference your response in 3a, as appropriate.

NA

B) Independent reviewers: If your protocol is ***More than Minimal Risk***, please list the names And contact information (telephone, e-mail, address) of 3 experts in your field who can independently evaluate your proposal and assist the IRB in the review process.

NA

Protocol Application checklist

☒ A one-paragraph abstract describing the protocol

☐ Copy of IRB approval from collaborative institutions (NA)

☒ Investigator(s) vita

☒ Consent documents

☒ Questionnaires, measures, survey instruments

☒ Advertisements/recruitment letters

Categories of Research That May Be Reviewed by the Institutional Review Board (IRB) through an Expedited Review Procedure

Please check the category that applies

☐**1)** Clinical studies of drugs and medical devices only when condition (a) or (b) is met. (a) Research on drugs for which an investigational new drug application (21 CFR Part 312) is not required. (Note: Research on marketed drugs that significantly increases the risks or decreases the acceptability of the risks associated with the use of the product is not eligible for expedited review.) (b) Research on medical devices for which (i) an investigational device exemption application (21 CFR Part 812) is not required; or (Soliman & Tuunainen) the medical device is cleared/approved for marketing and the medical device is being used in accordance with its cleared/approved labeling.	☐ **2)** Collection of blood samples by finger stick, heel stick, ear stick, or venipuncture as follows: (a) from healthy, nonpregnant adults who weigh at least 110 pounds. For these subjects, the amounts drawn may not exceed 550 ml in an 8 week period and collection may not occur more frequently than 2 times per week; or (b) from other adults and children, considering the age, weight, and health of the subjects, the collection procedure, the amount of blood to be collected, and the frequency with which it will be collected. For these subjects, the amount drawn may not exceed the lesser of 50 ml or 3 ml per kg in an 8 week period and collection may not occur more frequently than 2 times per week.
☐**3)** Prospective collection of biological specimens for research purposes by noninvasive means. Examples: (a) hair and nail clippings in a nondisfiguring manner; (b) deciduous teeth at time of exfoliation or if routine patient care indicates a need for extraction; (c) permanent teeth if routine patient care indicates a need for extraction; (d) excreta and external secretions (including sweat); (e) uncannulated saliva collected either in an unstimulated fashion or stimulated by chewing gumbase or wax or by applying a dilute citric solution to the tongue; (f) placenta removed at delivery; (g) amniotic fluid obtainedat the time of rupture of the membrane prior to or during labor; (h) supra- and subgingival dental plaque and calculus, provided the collection procedure is not more invasive than routine prophylactic scaling of the teeth and the process is accomplished in accordance with accepted prophylactic techniques; (i) mucosal and skin cells collected by buccal scraping or swab, skin swab, or mouth washings; (j) sputum collected after saline mist nebulization.	☐ **4)** Collection of data through noninvasive procedures (not involving general anesthesia or sedation) routinely employed in clinical practice, excluding procedures involving x-rays or microwaves. Where medical devices are employed, they must be cleared/approved for marketing. (Studies intended to evaluate the safety and effectiveness of the medical device are not generally eligible for expedited review, including studies of cleared medical devices for new indications.) Examples: (a) physical sensors that are applied either to the surface of the body or at a distance and do not involve input of significant amounts of energy into the subject or an invasion of the subject's privacy; (b) weighing or testing sensory acuity; (c) magnetic resonance imaging; (d) electrocardiography, electroencephalography, thermography, detection of naturally occurring radioactivity, electroretinography, ultrasound, diagnostic infrared imaging, doppler blood flow, and echocardiography; (e) moderate exercise, muscular strength testing, body composition assessment, and flexibility testing where appropriate given the age, weight, and health of the individual.
☐**5)** Research involving materials (data, documents, records, or specimens) that have been collected, or will be collected solely for nonresearch purposes (such as medical treatment or diagnosis). Some research in this category may be exempt from the HHS regulations for the protection of human subjects. 45 CFR 46.101(b(4). This listing refers only to research that is not exempt.)	☐ **6)** Collection of data from voice, video, digital, or image recordings made for research purposes.
☒**7)** Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies. Some research in this category may be exempt from the HHS regulations for the protection of human subjects. 45 CFR 46.101(b(Kaikkonen et al.) and (b)(3). This listing refers only to research that is not exempt.)	

Appendix 4. IRB Approval

UMBC

AN HONORS UNIVERSITY IN MARYLAND

Office of Research Protections and Compliance
University of Maryland, Baltimore County
1000 Hilltop Circle
Baltimore, MD 21250
PHONE: 410-455-2737
EMAIL: compliance@umbc.edu
WEB: research.umbc.edu

Date: June 7, 2016
To: Kristen Richards, Anthony F Norcio
Re: Expedited Review Approval
Protocol #: Y16AN12253

Your protocol entitled Risk Analysis of the Discoverability of Personal Data Survey has been **approved by expedited review** by the Institutional Review Board. This study fulfills the criteria for expedited review under 45 CFR 46.110, category # 7 as ☐ *less than minimal risk* or ☒ *minimal risk.*

Approval of this protocol will terminate on the below end date unless an Annual Continuation Report is submitted, in writing, to the IRB. The Office of Research Protections and Compliance will send you an email reminder prior to the end of the protocol; it is your responsibility, however, to assure that project activities are not conducted past the expiration date.

Reporting Calendar

Original approval date	Current end date	The next Annual Continuation Report is due by	Expect a reminder to renew by
6/6/2016	***6/5/2017***	**05/08/2017**	**04/24/2017**

Investigators are responsible for reporting ***in writing*** to the IRB any changes to the human subject research protocol, measures or in the informed consent documents. This includes changes to the research design or procedures that could introduce new or increased risks to human subjects and thereby change the nature of the research. In addition, you must report any adverse events or unanticipated problems to the IRB for review and approval. All correspondence and materials used in this protocol must reference the above IRB number.

Investigators are also reminded that all UMBC IRB approved consent forms will display an expiration date at the bottom of each page. Please check this date carefully each time an approved consent form is used, as using an expired form to consent participants is considered a substantial deviation from the Federal regulations governing research involving human subjects.

The investigator(s) identified above are required to retain an IRB protocol file, including a record of IRB-related activity, data summaries and consent forms. This file is to be made available for review for internal procedural (audit) monitoring.

Expedited review approved by:

Susan Sonnenschein, Ph.D.
IRB Chair

Appendix 5. IRB Modification Request

IRB #:Y16AN12253
Investigator(s): Kirsten Richards, Dr. Anthony F. Norcio
Protocol Title: Risk Assessment of the Discoverability of Personal Data Survey
Date modification is submitted: 9/29/2016
Original Approval: Expedited ☒ Full Board ☐

Select all appropriate sections that describes the modification**.** Attach copy of revised protocol application **highlighting in yellow and underline)** indicating where changes are required. **ALL modified documents must be submitted in Microsoft Word format**

☐Change in Procedures	☐ **Add** ☐ **Delete** ☐ **Modify**
☐Change in Principal Investigator	☐ **Add** ☐ **Delete** ☐ **Modify** *Indicate CITI training completion dates*
☐Change in Study Personnel	☐ **Add** ☐ **Delete** ☐ **Modify** *Indicate CITI training completion dates* Has there any changes to any previous conflict of interest disclosure, as described by UMBC's Conflict of Interest policies? ☐ Yes ☐No ☐ N/A
☒Change in Measures	☒ **Add** ☐ **Delete** ☐ **Modify** *Attach copy of revisions **highlighting in yellow and underline.***
☐Change in / Add Sponsored Funding	☐ **Add** ☐ **Delete** ☐ **Modify** *Attach copy of the components of grant applications and contract proposals related to human subjects use [e.g. the Human Subjects section]as well as any amendment if it involves changes or additions to sponsored funding (e.g. submission of JIT materials, confirms to an agency that a grant has IRB approval, or adds a new award to an existing protocol).*
☐Change in Recruitment/Advertising	☐ **Add** ☐ **Delete** ☐ **Modify** *Attach copy of revisions **highlighting in yellow and underline.***

☐**Change in Number of Participants and/or Participant Selection**	☐ **Increase by:** Click here to enter text. ☐ **Decrease by:** Click here to enter text. **Resulting total to be enrolled:** Click here to enter text. **Reason for Change:** Click here to enter text.
☐ **Consent Process Change and/or Change in Consent Documents**	☐ **Add** ☐ **Delete** ☐ **Modify** *Attach a copy of the current approved consent document and a copy of the proposed document with changes* ***highlighted in yellow and underlined***
☐**Change in Data Collection, Storage and Confidentiality**	☐ **Add** ☐ **Delete** ☐ **Modify**
☐**Change in Location**	☐ **Add** ☐ **Delete** ☐ **Modify**
☐**Other Change**	**Specify:** Click here to enter text.

Reanalysis of Risk

☒This modification **does not** increase the risks to participants in the approved protocol
☐This modification **does** increase the risks to participants in the approved protocol

Provide a narrative summary of all proposed modifications with a description of how the modifications affect research risks and benefits. Also describe any event or new data that precipitated the change.
Click here to enter text.

Electronically submit the protocol and any accompanying documents to irbsubmissions@umbc.edu.

By typing your name, email address and date, the investigator(s) certify they will abide by all UMBC IRB policies and procedures and understand that no research activities will be conducted with human participants prior to obtaining the required approvals. All changes must be submitted and approved by the IRB prior to their implementation

Investigator's Signature: Kirsten Richards. Email: kirsten_richards@yahoo.com. Date: 9/29/2016

Investigator's Signature: Click here to enter text. Email: Click here to enter text. Date: Click here to enter a date.

Faculty Advisor's Signature: Click here to enter text. Email: Click here to enter text. Date: Click here to enter a date.

IRB Action

Approval of changes/modifications by the IRB Chair/Date Requires Expedited Review? **Yes ☐ No☐**

Approval of administrative changes/modifications, Office for Research Protections and Compliance/Date (modification request form) –06/06/2016

Appendix 6. IRB Modification Approval

UMBC

AN HONORS UNIVERSITY IN MARYLAND

Office of Research Protections and Compliance
University of Maryland, Baltimore County
1000 Hilltop Circle
Baltimore, MD 21250
PHONE: 410-455-2737
EMAIL: compliance@umbc.edu
WEB: research.umbc.edu

Date: 10/14/2016
To: Kristen Richards, Anthony F Norcio
Re: Notice of Action Modification Approval
Protocol #: Y16AN12253

Original approval date: 6/6/2016

Modifications submitted: **9/29/2016**

Your request for approval of changes made to the documents for your protocol entitled Risk Analysis of the Discoverability of Personal Data Survey has been **approved** by the Chair of the Institutional Review Board. This research was previously reviewed and approved by the IRB, where no greater than minimal risks to participants and no additional risks were identified.

Note that all other conditions and investigator responsibilities outlined in the original approval letter are still in force.

Appendix 7. IRB Approved Source Participant Consent

Whom to Contact about this study:

Principal Investigator: Kirsten Richards, Dr. A. F. Norcio
Department: Information Systems
Telephone number: 410-455-3206

Risk Assessment of the Discoverability of Personal Data Survey

I. INTRODUCTION/PURPOSE:

I am being asked to participate in a research study. The purpose of this study is to discover whether personal data, such as mother's maiden name, may be found online. I am being asked to volunteer based on my age and gender fitting the desired demographic information for the study. My involvement in this study will begin when I agree to participate and will continue until 12/2016. Approximately 6 other persons will also be providing their personal data for study. Approximately 50 people will participate by searching for information online regarding the other research subjects.

II. PROCEDURES:

As a participant in this study, I will be asked to allow other individuals to look up specific personal data, including my mother's maiden name, my pet's names, my middle name, nick names, and my children's names, using online resources. I will be asked to provide my demographic data including name, age, gender, and city/state as well as a photograph. After data collection, I will be asked to assess the accuracy of the data collected by questionnaire. I will complete the questionnaire from my home, school, or other convenient location. My participation in this study will last for approximately one month and will require the submission of demographic data, permission for data collection, and completion of the questionnaire. I will provide my demographic data and then be contacted approximately one month later to review the accuracy of data collected online. All data will be collected from a location of my choosing and no on-site visit will be required.

III. RISKS AND BENEFITS:

I have been informed that participation in this study may involve the following risks: Individuals participating in the study will be searching for personal data about me online and therefore, other personal data may be discovered by other study participants. I have been advised to secure my personal accounts using passwords and secondary authentication mechanisms that are not reliant on my personal data. My passwords should be secure and I understand how to create a secure password. I have also been informed that my participation in this research will not benefit me personally, but will help provide information about data that is available online, enabling the design of better security systems in the future.

CONFIDENTIALITY:

Any information learned and collected from this study in which I might be identified will remain confidential and will be disclosed ONLY if I give permission. My name, photograph and city and state will be supplied to participants to identify me online, but the data collected, and my assessment of the data accuracy will remain confidential.

Only the investigator and members of the research team will have access to these records. If information learned from this study is published, I will not be identified by name. By signing this form, however, I allow the research study investigator to make my records available to the University of Maryland Baltimore County (UMBC) Institutional Review Board (IRB) and regulatory agencies as required to do so by law.

Consenting to participate in this research also indicates my agreement that all information collected from me individually may be used by current and future researchers in such a fashion that my personal identity will be protected. Such use will include sharing anonymous information with other researchers for checking the accuracy of study findings and for future approved research that has the potential for improving human knowledge.

IV. COMPENSATION/COSTS:

My participation in this study will involve no cost to me. I will not be compensated for my participation.

V. **CONTACTS AND QUESTIONS:**

The principal investigator(s), Kirsten Richards, has offered to and has answered any and all questions regarding my participation in this research study. If I have any further questions, I can contact Kirsten Richards at kirsten5@umbc.edu or Dr. Anthony Norcio at norcio@umbc.edu

If I have any questions about my rights as a participant in this research study, contact the Office of Research Protections and Compliance at (410) 455-2737 or
compliance@umbc.edu.

VI. VOLUNTARY PARTICIPATION

I have been informed that my participation in this research study is voluntary and that I am free to withdraw or discontinue participation at any time. I have been informed that data collected for this study will be retained by the investigator and analyzed even if I choose to withdraw from the research. If I do choose to withdraw, the investigator and I have discussed my withdrawal and the investigator may use my information up to the time I decide to withdraw.

I will be given a copy of this consent form to keep.

VII. SIGNATURE FOR CONSENT

The above-named investigator has answered my questions and I agree to be a research participant in this study.

Participant's Name: _____ Date: _____

Participant's Signature: _____Date: _____

Investigator's Signature: _____Date: _____

Approved by the Permitted for use
UMBC Institutional Review Board **From 06/06/2016**
IRB Protocol Y16AN12253 **To 06/05/2017**
UMBC ORPC: 4/25/2017 2:04 PM

Appendix 8. IRB Approved Seeker Participant Consent

Whom to Contact about this study:
Principal Investigator: Kirsten Richards, Dr. A. F. Norcio
Department: Information Systems
Telephone number: 410-455-3206

Risk Assessment of the Discoverability of Personal Data Survey

I. INTRODUCTION/PURPOSE:

I am being asked to participate in a research study. The purpose of this study is to discover whether personal data, such as mother's maiden name, may be found online. I am being asked to volunteer based on my availability and willingness to participant in the study. My involvement in this study will begin when I agree to participate and will continue until 12/2016. About 50 persons will be invited to participate.

II. PROCEDURES:

As a participant in this study, I will be asked to look up personal data regarding other individuals using online resources. I will not use paid services, illegal data sources or use information I find illegally. I have been informed that it is illegal to attempt to access accounts belonging to other individuals. I will be asked to complete the survey from my home, school, or other convenient location. My participation in this study will last for approximately two weeks, although I may complete the survey more quickly if I wish, and will require the completion of the survey. All data will be collected from a location of my choosing and no on-site visit will be required.

III. RISKS AND BENEFITS:

I have been informed that participation in this study may involve the following risks: fatigue associated with normal computer usage. I may rest as needed. I have also been informed that my participation in this research will not benefit me personally, but will help provide information about data that is available online, enabling the design of better security systems in the future.

CONFIDENTIALITY:

Any information learned and collected from this study in which I might be identified will remain confidential and will be disclosed ONLY if I give permission. I provide permission to supply my name to an instructor at UMBC if I am receiving extra credit in a class for participation in the research study. I understand that I will be collecting

information about other participants as a part of this study and agree not to reveal their personal information or use their personal information for any purpose other than the completion of the study. I will maintain the confidentiality of the personal information I learn online.

Only the investigator and members of the research team will have access to these records. If information learned from this study is published, I will not be identified by name. By signing this form, however, I allow the research study investigator to make my records available to the University of Maryland Baltimore County (UMBC) Institutional Review Board (IRB) and regulatory agencies as required to do so by law.

Consenting to participate in this research also indicates my agreement that all information collected from me individually may be used by current and future researchers in such a fashion that my personal identity will be protected. Such use will include sharing anonymous information with other researchers for checking the accuracy of study findings and for future approved research that has the potential for improving human knowledge.

IV. COMPENSATION/COSTS:

My participation in this study will involve no cost to me.
Subject to IRB and instructor approval, I may be offered extra credit at UMBC if I am a current student in a class offering extra credit for participation. I have been informed if I am eligible for extra credit for participating in research. My name will be provided to my instructor for the purpose of distributing extra credit, if applicable.

V. CONTACTS AND QUESTIONS:

The principal investigator, Kirsten Richards, has offered to and has answered any and all questions regarding my participation in this research study. If I have any further questions, I can contact Kirsten Richards at kirsten5@umbc.edu or Dr. Anthony Norcio at norcio@umbcedu

If I have any questions about my rights as a participant in this research study, contact the Office of Research Protections and Compliance at (410) 455-2737 or

compliance@umbc.edu.

VI. VOLUNTARY PARTICIPATION

I have been informed that my participation in this research study is voluntary and that I am free to withdraw or discontinue participation at any time. I have been informed that data collected for this study will be retained by the investigator and analyzed even if I

choose to withdraw from the research. If I do choose to withdraw, the investigator and I have discussed my withdrawal and the investigator may use my information up to the time I decide to withdraw.

I will be given a copy of this consent form to keep.

VII. SIGNATURE FOR CONSENT

The above-named investigator has answered my questions and I agree to be a research participant in this study.

Participant's Name: _____ Date: _____

Participant's Signature: _____Date: _____

Investigator's Signature: _____Date: _____

Approved by the
UMBC Institutional Review Board
IRB Protocol Y16AN12253

Permitted for use
From 06/06/2016
To 06/07/2017
UMBC ORPC: 4/25/2017 2:04 PM

Appendix 9. IRB Approved Source Participant Recruitment

Source Participant Recruitment:
How much information is available about you online? Current research at UMBC is exploring how much information can be discovered about individuals online. Participants in this research will have the opportunity to discover whether specific information about them can be found by other individuals using online resources. Participation in the study requires supplying the researchers with permission to allow other individuals to search for your information. We will need your name, city and state and possibly a current photograph. Your involvement will include providing your permission and verifying whether the data discovered in the research is accurate. If you are interested, please contact:

Kirsten Richards
Kirsten5@umbc.edu

Seeker Participant Recruitment:
Participants are needed for a research project exploring the availability of personal information online. To participate, you will complete a survey and discover whether you can find specific data about other individuals, such as their pet's names. If you are interested, please contact:

Kirsten Richards
Kirsten5@umbc.edu

Approved by the UMBC Institutional Review Board
Permitted for use From 06/06/2016 To 06/05/2017
IRB Protocol Y16AN12253

UMBC ORPC: 4/25/2017 2:04 PM

Appendix 10. Accuracy of *Mother's Maiden Name* for DK

	Frequency	Percent
Incorrect	31	52.5
Not found	28	47.5
Total	59	100.0

Appendix 11. Familiarity *Mother's Maiden Name* for DK

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Incorrect	31	3.29	1.371	.246	2.79	3.79	1	5
Not Found	26	3.00	1.166	.229	2.53	3.47	1	5
Total	57	3.16	1.279	.169	2.82	3.50	1	5

Appendix 12. Time of *Mother's Maiden* Name for DK

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Incorrect	30	34.10	27.566	5.033	23.81	44.39	5	120
Not found	25	23.36	13.826	2.765	17.65	29.07	6	60
Total	55	29.22	22.851	3.081	23.04	35.40	5	120

Appendix 13. Locations of *Mother's Maiden Name* for DK

Locations	Number
411	2
Ancestry	4
AOL	1
BCPS (Baltimore County Public School)	13
Been Verified	6
Birth Records (unspecified)	1
Bing	2
biography (unspecified)	1
City Freq	1

Funeral Home Guestbook	2
Facebook	44
Google	25
Instant Checkmake	1
LinkedIn	9
Montgomory county races	1
My Heritage	6
My Life	3
Obituary	3
Pipl	7
Phone Book (unspecified)	7
Reverse Phone (unspecified)	1
Social Media	3
Truth Finder	5
Twitter	2
White Pages	35
Wood Briedge Elementary School	4
Yahoo	1
Zoom Info	1

Appendix 14. Perceived Difficulty of *Mother's Maiden Name* by Difficulty Score for DK

	Frequency	**Percent**
1	37	62.7
2	13	22.0
3	6	10.2
5	3	5.1
Total	59	100.0

Appendix 15. Perceived Difficulty of *Mother's Maiden Name* by Accuracy for DK

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Incorrect	31	2.10	1.193	.214	1.66	2.53	1	5
Not Found	26	1.12	.431	.085	.94	1.29	1	3
Total	57	1.65	1.044	.138	1.37	1.93	1	5

Appendix 16. Accuracy of *Nickname* for DK

	Frequency	**Percent**
Missing	3	5.1
Correct	4	6.8
Incorrect	21	35.6
Not found	31	52.5
Total	59	100.0

Appendix 17. Familiarity Compared by Accuracy of *Nickname* for DK

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	4	3.25	2.062	1.031	-.03	6.53	1	5
Incorrect	21	3.38	1.359	.297	2.76	4.00	1	5
Not found	29	3.03	1.117	.208	2.61	3.46	1	5
Total	54	3.19	1.275	.173	2.84	3.53	1	5

Appendix

Appendix 18. Time Compared by Accuracy of *Nickname* for DK

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	4	16.25	2.500	1.250	12.27	20.23	15	20
Incorrect	20	19.95	17.563	3.927	11.73	28.17	2	65
Not found	29	16.90	10.352	1.922	12.96	20.83	5	45

Total	53	18.00	13.159	1.807	14.37	21.63	2	65

Appendix 19. Locations of *Nickname* for DK

Location	Number
411	0
Ancestry	1
AOL	0
BCPS (Baltimore County Public School)	1
Been Verified	0
Birth Records (unspecified)	0
Bing	0
biography (unspecified)	1
City Freq	0
Funeral Home Guestbook	1
Facebook	5
Google	5
Instant Checkmake	0
LinkedIn	1
Montgomery county races	0
My Heritage	0
My Life	0
Obituary	0
Pipl	1
Phone Book (unspecified)	2
Reverse Phone (unspecified)	0
Social Media	0

Truth Finder	1
Twitter	0
White Pages	2
Wood Bridge Elementary School	1
Yahoo	0
Zoom Info	0

Appendix 20. Perceived Difficulty of *Nickname* for DK

	Frequency	**Percent**
1	41	69.5
2	5	8.5
3	8	13.6
4	1	1.7
5	1	1.7
Total	56	94.9
Missing	3	5.1
Total	59	100.0

Appendix 21. Perceived Difficulty of *Nickname* by Accuracy for DK

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	4	2.00	1.155	.577	.16	3.84	1	3
Incorrect	21	2.05	1.161	.253	1.52	2.58	1	5
Not found	29	1.07	.371	.069	.93	1.21	1	3
Total	54	1.52	.947	.129	1.26	1.78	1	5

Appendix 22. Accuracy of *Children's Names* for DK

	Frequency	**Percent**
Missing	3	5.1
Correct/Incomplete	6	10.2
Incorrect	22	37.3
Not found	24	40.7
Partially Correct/Incomplete	4	6.8
Total	59	100.0

Appendix 23. Familiarity with Search for Children's Names for DK

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct/ Incomplete	6	3.83	.983	.401	2.80	4.87	2	5
Partially Correct/ Incomplete	4	3.25	.957	.479	1.73	4.77	2	4
Incorrect	22	2.95	1.397	.298	2.34	3.57	1	5
Not Found	24	3.21	1.250	.255	2.68	3.74	1	5
Total	56	3.18	1.266	.169	2.84	3.52	1	5

Appendix 24. Search time of *Children's Names* for DK

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct/Incomplete	6	18.33	16.931	6.912	.57	36.10	5	40
Partially Correct/ Incomplete	4	20.50	17.156	8.578	-6.80	47.80	5	45
Incorrect	21	17.67	14.871	3.245	10.90	24.44	4	65
Not Found	24	16.79	13.214	2.697	11.21	22.37	3	60
Total	55	17.56	14.147	1.908	13.74	21.39	3	65

Appendix 25. Locations of *Children's Names* for DK

Location	**Number**
411	0
Ancestry	0
AOL	0
BCPS (Baltimore County Public School)	0
Been Verified	1
Birth Records (unspecified)	0
Bing	0

biography (unspecified)	0
City Freq	0
Funeral Home Guestbook	0
Facebook	16
Google	4
Instant Checkmate	0
LinkedIn	2
Montgomery county races	1
My Heritage	2
My Life	1
Obituary	1
Pipl	1
Phone Book (unspecified)	0
Reverse Phone (unspecified)	0
Social Media	1
Truth Finder	1
Twitter	1
White Pages	4
Wood Bridge Elementary School	0
Yahoo	0
Zoom Info	0

Appendix 26. Perceived Difficulty of *Children's Names'* for DK

	Frequency	Percent
1	35	59.3
2	8	13.6

3	7	11.9
4	5	8.5
5	4	6.8
Total	59	100.0

Appendix 27. Perceived Difficulty of *Children's Names* by Accuracy for DK

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct/Incomplete	6	3.67	1.506	.615	2.09	5.25	1	5
Partially Correct/ Incomplete	4	2.75	1.500	.750	.36	5.14	2	5
Incorrect	22	2.32	1.249	.266	1.76	2.87	1	5
Not Found	24	1.04	.204	.042	.96	1.13	1	2
Total	6	1.95	1.506	.615	2.09	5.25	1	5

Appendix 28. Accuracy of *Pet's Names* for DK

	Frequency	Percent
Missing	11	18.6
Correct	20	33.9
Not found	28	47.5
Total	59	100.0

Appendix 29. Familiarity with search of *Pet's Names* for DK

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	20	3.00	1.338	.299	2.37	3.63	20	3.00
Not Found	28	3.25	1.295	.245	2.75	3.75	28	3.25
Total	48	3.15	1.304	.188	2.77	3.52	48	3.15

Appendix 30. Time of *Pet's Names* for DK

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	20	10.65	10.007	2.238	5.97	15.33	1	45
Not Found	28	17.89	14.743	2.786	12.18	23.61	5	60
Total	48	14.88	13.355	1.928	11.00	18.75	1	60

Appendix 31. Locations of *Pet's Names* for DK

Location	**Number**
411	0
Ancestry	0
AOL	0
BCPS (Baltimore County Public School)	0
Been Verified	0
Birth Records (unspecified)	0
Bing	1
biography (unspecified)	0
City Freq	0
Funeral Home Guestbook	0
Facebook	7
Google	5
Instant Checkmate	0
LinkedIn	2
Montgomery county races	0
My Heritage	1
My Life	0

Obituary	0
Pipl	1
Phone Book (unspecified)	0
Reverse Phone (unspecified)	0
Social Media	2
Truth Finder	0
Twitter	0
White Pages	1
Wood Bridge Elementary School	0
Yahoo	1
Zoom Info	0

Appendix 32. Frequency of Difficulty Ratings of *Pet's Names* for DK

	Frequency	**Percent**
1	41	69.5
2	1	1.7
3	2	3.4
4	3	5.1
5	2	3.4
6	1	1.7
Total	50	84.7
Missing	9	15.3
Total	59	100.0

Appendix 33. Perceived Difficulty of *Pet's Names* for DK

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	20	2.30	1.750	.391	1.48	3.12	1	6

Not Found	28	1.04	.189	.036	.96	1.11	1	2
Total	48	1.56	1.287	.186	1.19	1.94	1	6

Appendix 34. Accuracy of *Middle Name* for DK

	Frequency	Percent	Valid Percent	Cumulative Percent
Correct	35	68.6	68.6	68.6
Incorrect	7	13.7	13.7	82.4
Not found	9	17.6	17.6	100.0
Total	51	100.0	100.0	

Appendix 35. Familiarity by Accuracy of *Middle Name* for DK

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	20	3.30	1.261	.282	2.71	3.89	1	5
Correct/Incomplete	2	4.50	.707	.500	-1.85	10.85	4	5
Incorrect	13	2.62	1.261	.350	1.85	3.38	1	4
Not Found	19	3.32	1.157	.265	2.76	3.87	2	5
Total	54	3.19	1.245	.169	2.85	3.52	1	5

Appendix 36. Time by Accuracy of *Middle Name* for DK

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	20	10.45	10.405	2.327	5.58	15.32	1	45
Correct/Incomplete	2	11.00	5.657	4.000	39.82	61.82	7	15
Incorrect	13	18.31	20.698	5.741	5.80	30.82	1	65
Not found	19	20.11	17.282	3.965	11.78	28.43	2	60
Total	54	15.76	16.053	2.185	11.38	20.14	1	65

Appendix 37. Locations of *Middle Name* for DK

Locations	**Number**
411	1

Ancestry	1
AOL	0
BCPS (Baltimore County Public School)	2
Been Verified	1
Birth Records (unspecified)	0
Bing	0
biography (unspecified)	0
City Freq	0
Funeral Home Guestbook	0
Facebook	6
Google	3
Instant Checkmate	0
LinkedIn	1
Montgomery county races	0
My Heritage	2
My Life	0
Obituary	0
Pipl	2
Phone Book (unspecified)	0
Reverse Phone (unspecified)	0
Social Media	0
Truth Finder	1
Twitter	0

White Pages	18
Wood Bridge Elementary School	0
Yahoo	0
Zoom Info	0

Appendix 38. Perceived Difficulty of *Middle Name* for DK

	Frequency	Percent
1	23	39.0
2	12	20.3
3	5	8.5
4	9	15.3
5	4	6.8
6	3	5.1
Total	56	94.9

Appendix 39. Perceived Difficulty by Accuracy of *Middle Name* for DK

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	20	3.55	1.468	.328	2.86	4.24	2	6
Correct/Incomplete	2	2.50	.707	.500	-3.85	8.85	2	3
Incorrect	13	2.77	1.589	.441	1.81	3.73	1	5
Not Found	19	1.00	.000	.000	1.00	1.00	1	1
Total	54	2.43	1.609	.219	1.99	2.87	1	6

Appendix 40. Accuracy of *Mobile Phone Number* for DK

	Frequency	**Percent**
Missing	7	11.9
Incorrect	20	33.9
Not found	32	54.2
Total	59	100.0

Appendix 41. Familiarity by Accuracy of Mobile Phone Number for DK

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Incorrect	20	2.85	1.268	.284	2.26	3.44	1	5
Not found	32	3.25	1.270	.225	2.79	3.71	1	5
Total	52	3.10	1.272	.176	2.74	3.45	1	5

Appendix 42. Search Time for *Mobile Phone Number* for DK

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Incorrect	20	20.80	19.495	4.359	11.68	29.92	2	65
Not found	32	17.25	13.956	2.467	12.22	22.28	5	60
Total	52	18.62	16.218	2.249	14.10	23.13	2	65

Appendix 43. Locations of *Mobile Phone Number* for DK

Locations	**Number**
411	0
Ancestry	0
AOL	0
BCPS (Baltimore County Public School)	8
Been Verified	2
Birth Records (unspecified)	0
Bing	1
biography (unspecified)	0
City Freq	1
Funeral Home Guestbook	0
Facebook	2

Google	3
Instant Checkmate	0
LinkedIn	1
Montgomery county races	0
My Heritage	0
My Life	0
Obituary	0
Pipl	1
Phone Book (unspecified)	4
Reverse Phone (unspecified)	1
Social Media	0
Truth Finder	0
Twitter	0
White Pages	4
Wood Bridge Elementary School	2
Yahoo	0
Zoom Info	1

Appendix 44. Frequency of Difficulty of *Mobile Phone Number* for DK

	Frequency	**Percent**
1	41	69.5
2	7	11.9
3	5	8.5
4	2	3.4
5	1	1.7
6	1	1.7
Total	57	96.6
Missing	2	3.4

Total	59	100.0

Appendix 45. Perceived Difficulty of *Mobile Phone Number* for DK

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Incorrect	20	2.50	1.395	.312	1.85	3.15	1	6
Not found	32	1.06	.354	.063	.94	1.19	1	3
Total	52	1.62	1.140	.158	1.30	1.93	1	6

Appendix 46. DK Mother's Maiden Name ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	1.192	1	1.192	.725	.398
	Within Groups	90.387	55	1.643		
	Total	91.579	56			
Mother's Maiden Time	Between Groups	1572.922	1	1572.922	3.131	.083
	Within Groups	26624.460	53	502.348		
	Total	28197.382	54			
Mother Maiden Difficulty	Between Groups	13.619	1	13.619	15.815	.000
	Within Groups	47.364	55	.861		
	Total	60.982	56			

Appendix 47. DK Nickname ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	1.480	2	.740	.446	.643
	Within Groups	84.668	51	1.660		
	Total	86.148	53			
Nickname Time	Between Groups	123.610	2	61.805	.348	.708
	Within Groups	8880.390	50	177.608		
	Total	9004.000	52			
Nickname Difficulty	Between Groups	12.667	2	6.334	9.278	.000
	Within Groups	34.814	51	.683		
	Total	47.481	53			

Appendix 48. DK Children's Names

DK Children's Names ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	3.718	3	1.239	.763	.520
	Within Groups	84.496	52	1.625		
	Total	88.214	55			
Children's Time	Between Groups	52.569	3	17.523	.083	.969
	Within Groups	10754.958	51	210.882		
	Total	10807.527	54			
Children's Difficulty	Between Groups	43.025	3	14.342	14.393	.000
	Within Groups	51.814	52	.996		
	Total	94.839	55			

Appendix 49. DK Pet's Names ANOVA

DK Pet's Names ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	.729	1	.729	.423	.519
	Within Groups	79.250	46	1.723		
	Total	79.979	47			
Pet's Time	Between Groups	612.021	1	612.021	3.623	.063
	Within Groups	7771.229	46	168.940		
	Total	8383.250	47			
Pet's Difficulty	Between Groups	18.648	1	18.648	14.499	.000
	Within Groups	59.164	46	1.286		
	Total	77.813	47			

Appendix 50. DK Middle Name ANOVA

DK Middle Name ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	8.266	3	2.755	1.865	.148
	Within Groups	73.882	50	1.478		
	Total	82.148	53			
Middle Name Time	Between Groups	1052.362	3	350.787	1.391	.256
	Within Groups	12605.509	50	252.110		
	Total	13657.870	53			
Middle Name Difficulty	Between Groups	65.446	3	21.815	15.201	.000
	Within Groups	71.758	50	1.435		
	Total	137.204	53			

Appendix 51. DK Mobile Phone Number ANOVA

DK Mobile Phone Number ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	1.969	1	1.969	1.222	.274
	Within Groups	80.550	50	1.611		
	Total	82.519	51			
Mobile Phone Time	Between Groups	155.108	1	155.108	.585	.448
	Within Groups	13259.200	50	265.184		
	Total	13414.308	51			
Mobile Phone Difficulty	Between Groups	25.433	1	25.433	31.110	.000
	Within Groups	40.875	50	.818		
	Total	66.308	51			

GC Results

Appendix 52. Accuracy of Mother's Maiden Name for GC

	Frequency	Percent
Missing	4	7.8
Correct	1	2.0
Incorrect	32	62.7
Not found	14	27.5
Total	51	100.0

Appendix 53. Familiarity Compared by Accuracy of Mother's Maiden Name for GC

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	1	3.00	.	.	.	.	3	3
Incorrect	32	3.00	1.164	.206	2.58	3.42	1	5
Not Found	14	2.86	1.167	.312	2.18	3.53	1	5
Total	47	2.96	1.141	.166	2.62	3.29	1	5

Appendix 54. Time Compared by Accuracy of Mother's Maiden Name for GC

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	1	15.00	.	.	.	.	15	15
Incorrect	32	22.44	23.523	4.158	13.96	30.92	1	115
Not found	14	21.86	14.368	3.840	13.56	30.15	5	60
Total	47	22.11	20.795	3.033	16.00	28.21	1	115

Appendix 55. Locations of Mother's Maiden Name for GC

Location Total Reported	Number
Ancestry	2
AOL	1
BeenVerified	0
Birth Records	1
Facebook	18
GC's Dressage Website	0
Google	3
Moms.mn.gov	1
Mylife	1
Nuwber	1
Ohio Resident DB	3
Ohio Voters	1
Peoplesmart	0

Phone	1
Quanki	0
Spokeo	1
Truthfinder	4
Twitter	1
VoterRecords	0
White Pages	3

Appendix 56. Perceived Difficulty of Mother's Maiden Name by Difficulty Score for GC

	Frequency	**Percent**
1	18	35.3
2	8	15.7
3	10	19.6
4	6	11.8
5	5	9.8
Total	47	92.2
Missing	4	7.8
Total	51	100.0

Appendix 57. Perceived Difficulty of Mother's Maiden Name by Correctness for GC

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	1	4.00	.	.	.	.	4	4
Incorrect	32	2.97	1.257	.222	2.52	3.42	1	5
Not Found	14	1.00	.000	.000	1.00	1.00	1	1
Total	47	2.40	1.393	.203	2.00	2.81	1	5

Appendix 58. Accuracy of Nickname for GC

	Frequency	**Percent**
Missing	2	3.9
Correct	4	7.8
Incorrect	21	41.2
Not found	24	47.1
Total	51	

Appendix 59. Familiarity Compared by Accuracy of *Nickname* for GC

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	4	4.25	0.957	0.479	2.73	5.77	3	5
Incorrect	21	2.52	1.167	0.255	1.99	3.06	1	5
Not found	24	3.13	0.947	0.193	2.73	3.52	1	5
Total	49	2.96	1.136	0.162	2.63	3.29	1	5

4.1.3.2.3 Search Time of *Nickname* for GC

Appendix 60. Familiarity Time Compared by Accuracy of Nickname for GC

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	4	8.75	2.5	1.25	4.77	12.73	5	10
Incorrect	21	15.95	14.634	3.193	9.29	22.61	1	60
Not found	24	15.13	11.961	2.442	10.07	20.18	3	60
Total	49	14.96	12.721	1.817	11.31	18.61	1	60

Appendix 61. Locations of Nickname for GC

Location Total Reported	**Number of Instances**
Ancestry	0
AOL	0
BeenVerified	0
Birth Records	0
Facebook	23
Hannah Bigg's Dressage Website	0
Google	3
mn.gov	0
mylife	0
Nuwber	1
Ohio Resident DB	1
Ohio Voters	0
Peoplesmart	0
Phone	0
Quanki	0
Spokeo	0
Truthfinder	2
Twitter	0

VoterRecords	0
White Pages	1

Appendix 62. Perceived Difficulty of Nickname for GC

	Frequency	**Percent**
1	27	52.9
2	3	5.9
3	10	19.6
4	5	9.8
5	2	3.9
6	2	3.9
Total	49	96.1
Missing	2	3.9
Total	51	100.0

Appendix 63. Perceived Difficulty of *Nickname* by Accuracy for GC

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	4	3.00	2.160	1.080	-.44	6.44	1	6
Incorrect	21	3.29	1.231	.269	2.73	3.85	1	6
Not found	24	1.00	.000	.000	1.00	1.00	1	1
Total	49	2.14	1.486	.212	1.72	2.57	1	6

Appendix 64. Accuracy of Children's Names for GC

	Frequency	**Percent**
Missing	1	2.0
Correct/Incomplete	20	39.2
Incorrect	8	15.7
Not found	8	15.7
Partially Correct/Incomplete	14	27.5
Total	51	100.0

Appendix 65. Familiarity with Search by Accuracy group for GC

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct/ Incomplete	20	3.30	1.129	.252	2.77	3.83	1	5
Partially Correct/ Incomplete	14	2.57	.938	.251	2.03	3.11	1	4
Incorrect	8	2.75	1.389	.491	1.59	3.91	1	5

Not Found	8	2.50	1.195	.423	1.50	3.50	1	4
Total	50	2.88	1.154	.163	2.55	3.21	1	5

Appendix 66. Search time of Children's Names for GC

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct/ Incomplete	20	9.40	8.127	1.817	5.60	13.20	1	30
Partially Correct/ Incomplete	14	16.14	12.247	3.273	9.07	23.21	3	50
Incorrect	7	12.43	11.028	4.168	2.23	22.63	2	30
Not Found	8	20.88	17.900	6.329	5.91	35.84	5	60
Total	49	13.63	12.122	1.732	10.15	17.11	1	60

Appendix 67. Locations of Children's Names for GC

Location Total Reported	**Number of Instances**
Ancestry	0
AOL	0
BeenVerified	0
Birth Records	0
Facebook	37
Hannah Bigg's Dressage Website	0
Google	2
Mom.mn.gov	0
Mylife	0
Nuwber	1
Ohio Resident DB	2
Ohio Voters	0
Peoplesmart	0
Phone	0
Quanki	0
Spokeo	0
Truthfinder	1
Twitter	0
VoterRecords	0
White Pages	0

Appendix 68. Perceived Difficulty of Children's Names' for GC

	Frequency	Percent
1	9	17.6
2	4	7.8
3	7	13.7
4	14	27.5
5	10	19.6

6	5	9.8
Total	49	96.1
Missing	2	3.9
Total	51	100.0

Appendix 69. Perceived Difficulty of Children's Names by Accuracy for GC

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct/Incomplete	20	4.10	1.483	.332	3.41	4.79	1	6
Partially Correct/ Incomplete	14	3.71	.994	.266	3.14	4.29	2	5
Incorrect	7	4.14	1.574	.595	2.69	5.60	1	6
Not Found	8	1.38	1.061	.375	.49	2.26	1	4
Total	49	3.55	1.608	.230	3.09	4.01	1	6

Appendix 70. Accuracy of Pet's Names for GC

	Frequency	Percent
Missing	2	3.9
Correct	15	29.4
Incorrect	25	49.0
Not found	9	17.6
Total	51	100.0

Appendix 71. Familiarity with search of Pet's Names for GC

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	15	2.73	1.100	.284	2.12	3.34	1	5
Incorrect	25	2.96	1.241	.248	2.45	3.47	1	5
Not Found	9	2.89	1.167	.389	1.99	3.79	1	5
Total	49	2.88	1.166	.167	2.54	3.21	1	5

Appendix 72. Time of Pet's Names for GC

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	15	14.6	14.51	3.746	6.56	22.64	0	60
Incorrect	25	9.8	12.39	2.478	4.69	14.91	1	60

Not Found	9	11.33	5.196	1.732	7.34	15.33	5	20
Total	49	11.55	12.131	1.733	8.07	15.04	0	60

Appendix 73. Locations of Pet's Names for GC

Location Total Reported	Number of Instances
Ancestry	0
AOL	0
BeenVerified	0
Birth Records	0
Facebook	31
Hannah Bigg's Dressage Website	0
Google	1
.Gov	0
mylife.com	0
Nuwber	0
Ohio Resident DB	0
Ohio Voters	0
Peoplesmart.com	0
Phone	0
Quanki	0
Spokeo	0
Truthfinder	1
Twitter	1
VoterRecords.com	1
White Pages	1

Appendix 74. Difficulty by Frequency of Difficulty Selection of Pet's Names for GC

	Frequency	Percent
1	16	31.4
2	4	7.8
3	7	13.7
4	14	27.5
5	4	7.8
6	5	9.8
Total	50	98.0
Missing	1	2.0
Total	51	100.0

Appendix 75. Perceived Difficulty of Pet's Names for GC

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	15	2.47	1.552	0.401	1.61	3.33	1	6
Incorrect	25	4.04	1.306	0.261	3.5	4.58	1	6
Not Found	9	1.33	1	0.333	0.56	2.1	1	4
Total	49	3.06	1.701	0.243	2.57	3.55	1	6

Appendix 76. Accuracy of Middle Name for GC

	Frequency	Percent
Correct	35	68.6
Incorrect	7	13.7
Not found	9	17.6
Total	51	100.0

Appendix 77. Familiarity by Accuracy of Middle Name for GC

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	35	2.97	1.043	0.176	2.61	3.33	1	5
Incorrect	7	3.14	1.676	0.634	1.59	4.69	1	5
Not Found	8	2.5	1.195	0.423	1.5	3.5	1	4
Total	50	2.92	1.158	0.164	2.59	3.25	1	5

Appendix 78. Time by Accuracy of Middle Name for GC.

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	35	8.03	11.344	1.917	4.13	11.93	1	60
Incorrect	7	5.14	3.237	1.223	2.15	8.14	2	12
Not found	7	14	11.676	4.413	3.2	24.8	2	35
Total	49	8.47	10.757	1.537	5.38	11.56	1	60

Appendix 79. Locations of Middle Name for GC

Location Total Reported	**Number**
Ancestry	0
AOL	0
BeenVerified	3
Birth Records	1
Facebook	5

GC's Dressage Website	0
Google	6
.Gov	0
mylife	2
Nuwber	7
Ohio Resident DB	11
Ohio Voters	1
Peoplesmart	1
Phone	0
Quanki	1
Spokeo	0
Truthfinder	2
Twitter	0
VoterRecords.com	1
White Pages	5

Appendix 80. Perceived Difficulty of Middle Name for GC

	Frequency	Percent
1	9	17.6
3	6	11.8
4	9	17.6
5	18	35.3
6	9	17.6
Total	51	100.0

Appendix 81. Perceived Difficulty by Accuracy of Middle Name for GC

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	35	4.66	1.162	.196	4.26	5.06	1	6
Incorrect	7	4.43	.976	.369	3.53	5.33	3	6
Not Found	8	1.50	1.414	.500	.32	2.68	1	5
Total	50	4.12	1.637	.231	3.65	4.59	1	6

Appendix 82. Accuracy of Mobile Phone Number for GC

	Frequency	Percent
Correct/Incomplete	1	2.0
Incorrect	7	13.7
Not found	37	72.5
Total	45	88.2

Missing	6	11.8
Total	51	100.0

Appendix 83. Familiarity by Accuracy of Mobile Phone Number for GC

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct/Incomplete	1	2	.	.	.	.	2	2
Incorrect	7	2.14	0.9	0.34	1.31	2.97	1	3
Not found	37	3.11	1.1	0.181	2.74	3.47	1	5

Appendix 84. Locations of Mobile Phone Number for GC

Locations	**Number**
Ancestry	0
AOL	1
BeenVerified	0
Birth Records	0
Facebook	5
GC's Dressage Website	2
Google	4
MN.gov	0
Mylife	0
Nuwber	3
Ohio Resident DB	0
Ohio Voters	0
Peoplesmart	0
Phone	1
Quanki	0
Spokeo	0
Truthfinder	2
Twitter	0
VoterRecords	0
White Pages	4

Appendix 85. Perceived Difficulty of Mobile Phone by Difficulty Score for GC

	Frequency	Percent
1	39	76.5
2	1	2.0
3	6	11.8
5	2	3.9
6	2	3.9
Total	50	98.0

Missing	1	2.0
Total	51	100.0

Appendix 86. Perceived Difficulty of Mobile Phone Number for GC

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct/Incomplete	1	1	.	.	.	.	1	1
Incorrect	7	3.29	1.38	0.522	2.01	4.56	1	5
Not found	37	1.41	1.212	0.199	1	1.81	1	6
Total	45	1.69	1.395	0.208	1.27	2.11	1	6

ANOVA for GC

Appendix 87. GC Mother's Maiden Name ANVOA

GC Mother's Maiden Name ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	.201	2	.100	.074	.929
	Within Groups	59.714	44	1.357		
	Total	59.915	46			
Mother's Time	Between Groups	54.879	2	27.439	.061	.941
	Within Groups	19837.589	44	450.854		
	Total	19892.468	46			
Mother's Difficulty	Between Groups	40.350	2	20.175	18.128	.000
	Within Groups	48.969	44	1.113		
	Total	89.319	46			

Appendix 88. GC Nickname ANVOA

GC Nickname ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	11.305	2	5.653	5.137	.010
	Within Groups	50.613	46	1.100		
	Total	61.918	48			

Nickname Time	Between Groups	175.591	2	87.795	.532	.591
	Within Groups	7592.327	46	165.051		
	Total	7767.918	48			
Nickname Difficulty	Between Groups	61.714	2	30.857	32.052	.000
	Within Groups	44.286	46	.963		
	Total	106.000	48			

Appendix 89. GC Children's Name ANOVA

GC Children's Names ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	6.151	3	2.050	1.595	.203
	Within Groups	59.129	46	1.285		
	Total	65.280	49			
Children's Time	Between Groups	876.284	3	292.095	2.128	.110
	Within Groups	6177.104	45	137.269		
	Total	7053.388	48			
Children's Difficulty	Between Groups	46.733	3	15.578	9.058	.000
	Within Groups	77.389	45	1.720		
	Total	124.122	48			

Appendix 90. GC Pet's Names ANOVA

GC Pet's Names ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	.483	2	.242	.172	.843
	Within Groups	64.782	46	1.408		
	Total	65.265	48			
Pet's Time	Between Groups	216.522	2	108.261	.727	.489
	Within Groups	6847.600	46	148.861		
	Total	7064.122	48			
Pet's Difficulty	Between Groups	56.123	2	28.061	15.610	.000
	Within Groups	82.693	46	1.798		
	Total	138.816	48			

Appendix 91. GC Middle Name ANOVA

GC Middle Name ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	1.851	2	.926	.682	.511
	Within Groups	63.829	47	1.358		
	Total	65.680	49			
Middle Time	Between Groups	298.376	2	149.188	1.306	.281
	Within Groups	5255.829	46	114.257		
	Total	5554.204	48			

Middle Difficulty	Between Groups	65.680	2	32.840	23.529	.000
	Within Groups	65.600	47	1.396		
	Total	131.280	49			

Appendix 92. GC Mobile Phone Number ANOVA

GC Mobile Phone ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	6.375	2	3.188	2.765	.074
	Within Groups	48.425	42	1.153		
	Total	54.800	44			
Phone Difficulty	Between Groups	21.297	2	10.648	6.950	.002
	Within Groups	64.347	42	1.532		
	Total	85.644	44			

Results for KT

Appendix 93. Accuracy of Mother's Maiden Name for KT

	Frequency	Percent
Missing	3	4.8
Correct	2	3.2
Incorrect	19	30.6
Not found	38	61.3
Total	62	100.0

Appendix 94. Familiarity Compared by Accuracy for KT

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	2	3.50	.707	.500	-2.85	9.85	3	4
Incorrect	19	3.42	1.261	.289	2.81	4.03	1	5
Not Found	38	3.08	1.100	.178	2.72	3.44	1	5
Total	59	3.20	1.141	.149	2.91	3.50	1	5

Appendix 95. Time Compared by Accuracy for KT

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	2	57.50	3.536	2.500	25.73	89.27	55	60
Incorrect	19	20.47	17.105	3.924	12.23	28.72	1	60
Not found	38	22.11	20.098	3.260	15.50	28.71	4	120
Total	59	22.78	19.807	2.579	17.62	27.94	1	120

Appendix 96. Locations of Mother's Maiden Name for KT

Ancestry	0
AOL	0
Been Verified	1
California Public records	1
Church Website	0
dbcomp.co	0
Desert Christian school	0
Life Pacific College	4
Family Search	1
Facebook	7
Google	7

Instant Checkmate	1
Intelius	0
Mylife	0
Nuwber	1
Lancaster People	0
LinkedIn	2
Obituary	3
Peoples Smart	0
Pipl	1
Phone Book (unspecified)	0
Public Record 360	0
Quanki	1
Radaris	1
Rate My Professor	0
Spokeo	0
Truth Finder	1
Twitter	0
Wikipedia	2
White Pages	5
Yellow Pages	0
Youtube	0
Zaba Search	0

Appendix 97. Perceived Difficulty of Mother's Maiden Name by Difficulty Score for KT

	Frequency	**Percent**
1	38	61.3
2	7	11.3
3	5	8.1
4	6	9.7
5	4	6.5
Total	60	96.8
Missing	2	3.2
Total	62	100

Appendix 98. Perceived Difficulty of *Mother's Maiden Name* by Correctness for KT

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	2	2.50	.707	.500	-3.85	8.85	2	3
Incorrect	19	3.32	1.250	.287	2.71	3.92	1	5

Not Found	38	1.11	.509	.083	.94	1.27	1	4
Total	59	1.86	1.319	.172	1.52	2.21	1	5

Appendix 99. Accuracy of Nickname for KT

	Frequency	**Percent**
Missing	11	17.7
Correct	10	16.1
Incorrect	25	40.3
Not found	13	21.0
Partially Correct/ Incomplete	3	4.8
Total	62	100.0

Appendix 100. Familiarity Compared by Accuracy of Nickname for KT

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	10	3.90	.876	.277	3.27	4.53	3	5
Partially Correct/Incomplete	3	3.00	2.000	1.155	-1.97	7.97	1	5
Incorrect	25	2.92	1.187	.237	2.43	3.41	1	5
Not found	18	3.11	1.079	.254	2.57	3.65	1	5
Total	56	3.16	1.172	.157	2.85	3.47	1	5

Appendix 101. Familiarity Time Compared by Accuracy of Nickname for KT

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	10	13.40	9.834	3.110	6.37	20.43	4	30
Partially Correct/Incomplete	3	12.00	11.533	6.658	-16.65	40.65	3	25
Incorrect	25	13.32	10.351	2.070	9.05	17.59	5	45
Not found	18	16.67	10.738	2.531	11.33	22.01	3	45
Total	56	14.34	10.284	1.374	11.59	17.09	3	45

Appendix 102. Locations of Nickname for KT

Location	Number
Ancestry	0
AOL	0
BeenVerified	1
CA Public records	3
Church	0
dbcomp.co	0
Desert Christian school	0
Life Pacific College	10
Familysearch	0
Facebook	14
Google	4
Instantcheckmate	0
Intelius	0
Mylife	0
Nuwber	0
Lancasterpeople	0
LinkedIn	0
Obituary	0
Peoplesmart.com	0
Pipl	0
Phone	0
PublicRecords360	3
Quanki	0
Radaris	0
Ratemyprofessor	4
Spokeo	0
Truthfinder	1
Twitter	0
Wikipedia	0

White Pages	0
Yellow Pages	0
Youtube	0
Zabasearch	0

Appendix 103. Perceived Difficulty of Nickname for KT

	Frequency	**Percent**
1	24	38.7
2	6	9.7
3	9	14.5
4	10	16.1
5	7	11.3
6	2	3.2
Total	58	93.5
System	4	6.5
Total	62	100.0

Appendix 104. Perceived Difficulty of Nickname by Accuracy for KT

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	10	4.10	.994	.314	3.39	4.81	3	6
Partially Correct/Incomplete	3	3.67	1.155	.667	.80	6.54	3	5
Incorrect	25	3.12	1.481	.296	2.51	3.73	1	6
Not found	18	1.00	.000	.000	1.00	1.00	1	1
Total	56	2.64	1.612	.215	2.21	3.07	1	6

Appendix 105. Accuracy of Children's Names for KT

	Frequency	**Percent**
Missing	13	21.0
Correct	1	1.6

Correct/Incomplete	13	21.0
Partially Correct/Incomplete	21	33.9
Incorrect	11	17.7
Not found	3	4.8
Total	62	100

Appendix 106. Familiarity with Search of Children's Names for KT

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	1	3.00	.	.	.	.	3	3
Correct/ Incomplete	13	3.00	1.414	.392	2.15	3.85	1	5
Partially Correct/ Incomplete	21	3.10	1.091	.238	2.60	3.59	1	5
Incorrect	11	2.82	1.250	.377	1.98	3.66	1	5
Not Found	12	3.67	.888	.256	3.10	4.23	2	5
Total	58	3.14	1.161	.153	2.83	3.44	1	5

Appendix 107. Search time of Children's Names for KT

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	1	5.00	.	.	.	.	5	5
Correct/ Incomplete	13	18.15	12.562	3.484	10.56	25.75	3	45
Partially Correct/ Incomplete	20	15.00	16.147	3.611	7.44	22.56	0	60
Incorrect	11	15.18	10.216	3.080	8.32	22.04	2	35
Not Found	12	11.67	6.140	1.772	7.77	15.57	1	20
Total	57	14.88	12.441	1.648	11.58	18.18	0	60

Appendix 108. Locations of Children's Names for KT

Locations	**Number**
Ancestry	1

AOL	0
Been Verified	4
California Public records	1
Church	0
dbcomp.co	0
Desert Christian school	0
Life Pacific College	4
Family Search	0
Facebook	15
Google	3
Instant Checkmate	0
Intelius	2
Mylife	1
Nuwber	5
Lancaster People	0
LinkedIn	2
Obituary	2
People Smart	1
Pipl	0
Phone	1
Public Records 360	1
Quanki	2
Radaris	0
Rate My Professor	0
Spokeo	1
Truthfinder	3
Twitter	0
Wikipedia	0
White Pages	4
Yellow Pages	0
Youtube	0

Zabasearch	0

Appendix 109. Perceived Difficulty of Children's Names' for KT

	Frequency	Percent
1	20	32.3
2	11	17.7
3	6	9.7
4	12	19.4
5	9	14.5
6	3	4.8
Total	61	98.4
Missing	1	1.6
Total	62	100

Appendix 110. Perceived Difficulty of Children's Names by Accuracy for KT

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	1	3.00	.	.	.	.	3	3
Correct/Incomplete	13	3.54	1.330	.369	2.73	4.34	2	6
Partially Correct/Incomplete	21	4.05	1.284	.280	3.46	4.63	2	6
Incorrect	11	1.91	1.221	.368	1.09	2.73	1	4
Not Found	12	1.08	.289	.083	.90	1.27	1	2
Total	58	2.90	1.640	.215	2.47	3.33	1	6

Appendix 111. Accuracy of Pet's Names for KT

	Frequency	Percent
Missing	27	43.5
Correct	1	1.6
Incorrect	25	40.3
Not found	9	14.5
Total	62	100.0

Appendix 112. Familiarity with search of Pet's Names for KT

	N	Mean	SD	SE	95% Confidence Interval for Mean	Min	Max

					Lower Bound	Upper Bound		
Correct	1	5.00	.	.	.	.	5	5
Incorrect	25	3.00	1.155	.231	2.52	3.48	1	5
Not Found	22	3.05	1.327	.283	2.46	3.63	1	5
Total	48	3.06	1.245	.180	2.70	3.42	1	5

Appendix 113. Time of Pet's Names for KT

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	1	3.00	.	.	.	.	3	3
Incorrect	21	16.19	13.920	3.038	9.85	22.53	2	50
Not Found	22	16.91	11.380	2.426	11.86	21.95	4	60
Total	44	16.25	12.557	1.893	12.43	20.07	2	60

Appendix 114. Locations of Pet's Names for KT

Location	**Number**
Ancestry	0
AOL	0
Been Verified	0
California Public records	0
Church	0
dbcomp.co	0
Desert Christian school	0
Life Pacific College	1
Family Search	0
Facebook	11
Google	7
Instant Checkmate	0
Intelius	1
Mylife	0
Nuwber	0

Lancaster People	0
LinkedIn	1
Obituary	0
People Smart	0
Pipl	0
Phone	0
Public Records 360	0
Quanki	0
Radaris	0
Rate My Professor	0
Spokeo	0
Truth Finder	0
Twitter	1
Wikipedia	0
White Pages	2
Yellow Pages	0
Youtube	0
Zabasearch	0

Appendix 115. Familiarity with Search of Pets for KT

	Frequency	**Percent**
1	39	62.9
2	7	11.3
3	1	1.6
4	6	9.7
5	2	3.2
6	2	3.2
Total	57	91.9
Missing	5	8.1
Total	62	100.0

Appendix 116. Perceived Difficulty of Pet's Names for KT

	N	Mean	SD	SE	95% Confidence Interval for Mean	Min	Max

					Lower Bound	Upper Bound		
Correct	1	6.00	.	.	.	.	6	6
Incorrect	25	2.60	1.555	.311	1.96	3.24	1	6
Not Found	22	1.00	.000	.000	1.00	1.00	1	1
Total	48	1.94	1.493	.216	1.50	2.37	1	6

Appendix 117. Accuracy of Middle Name for KT

	Frequency	Percent
Missing	6	9.7
Correct	33	53.2
Correct/Incomplete	9	14.5
Incorrect	10	16.1
Not found	4	6.5
Total	62	100.0

Appendix 118. Familiarity by Accuracy of Middle Name for KT

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	33	3.18	1.185	.206	2.76	3.60	1	5
Correct/Incomplete	9	2.78	1.093	.364	1.94	3.62	1	5
Incorrect	10	2.80	1.317	.416	1.86	3.74	1	5
Not Found	7	3.57	.787	.297	2.84	4.30	3	5
Total	59	3.10	1.155	.150	2.80	3.40	1	5

Appendix 119. Time by Accuracy of Middle Name for KT

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	31	12.35	13.235	2.377	7.50	17.21	0	50
Correct/Incomplete	9	8.56	6.064	2.021	3.89	13.22	2	20
Incorrect	10	13.70	11.245	3.556	5.66	21.74	5	39
Not Found	7	12.00	10.328	3.904	2.45	21.55	1	30
Total	57	11.95	11.547	1.529	8.88	15.01	0	50

Appendix 120. Locations of Middle Name for KT

Locations	Number
Ancestry	0
AOL	0
Been Verified	3
California Public Records	3
Church Website	0
dbcomp.co	0
Desert Christian school	0
Life Pacific College	1
Family Search	0
Facebook	10
Google	5
Instant Checkmate	0
Intelius	0
Mylife	2
Nuwber	8
Lancaster People	0
LinkedIn	1
Obituary	0
People Smart	0
Pipl	0
Phone Book (not specified)	0
Public Records 360	2
Quanki	0
Radaris	0
Ratemyprofessor	0
Spokeo	2
Truth Finder	2

Twitter	1
Wikipedia	0
White Pages	14
Yellow Pages	1
Youtube	2
Zaba search	0

Appendix 121. Perceived Difficulty of Middle Name for KT

	Frequency	**Percent**
1	11	17.7
2	9	14.5
3	11	17.7
4	8	12.9
5	17	27.4
6	6	9.7
Total	62	100.0

Appendix 122. Perceived Difficulty by Accuracy of Middle Name for KT

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	33	4.33	1.339	.233	3.86	4.81	2	6
Correct/Incomplete	9	3.78	1.093	.364	2.94	4.62	2	6
Incorrect	10	2.50	1.354	.428	1.53	3.47	1	5
Not Found	7	1.43	.787	.297	.70	2.16	1	3
Total	59	3.59	1.609	.210	3.17	4.01	1	6

Appendix 123. Accuracy of Mobile Phone Number for KT

	Frequency	**Percent**
Missing	8	12.9
Correct	4	6.5
Incorrect	41	66.1
Not found	9	14.5

Total	62	100.0

Appendix 124. Familiarity by Accuracy of Mobile Phone Number for KT

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	4	2.75	1.708	.854	.03	5.47	1	5
Incorrect	41	3.07	1.104	.172	2.72	3.42	1	5
Not Found	9	3.33	1.500	.500	2.18	4.49	1	5
Total	54	3.09	1.202	.164	2.76	3.42	1	5

Appendix 125. Time of Mobile Phone Number for KT

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	4	26.50	24.610	12.305	-12.66	65.66	1	60
Incorrect	41	11.39	11.369	1.775	7.80	14.98	1	45
Not Found	8	13.13	5.303	1.875	8.69	17.56	5	20
Total	53	12.79	12.416	1.706	9.37	16.21	1	60

Appendix 126. Locations of Mobile Phone Number for KT

Ancestry	0
AOL	0
Been Verified	9
CA Public records	1
Church Website	1
dbcomp.co	1
Desert Christian school	1
Life Pacific College	2
Family Search	0
Facebook	1
Google	3
Instant Checkmate	0

Intelius	0
My Life	0
Nuwber	11
Lancaster People	2
LinkedIn	1
Obituary	0
People Smart	0
Pipl	0
Phone	2
Public Records 360	1
Quanki	1
Radaris	1
Rate My Professor	0
Spokeo	1
Truthfinder	0
Twitter	0
Wikipedia	0
White Pages	5
Yellow Pages	1
Youtube	4
Zaba Search	1

Appendix 127. Difficulty Ratings of Mobile Phone Number for KT

	Frequency	Percent
1	17	27.4
2	4	6.5
3	13	21.0
4	8	12.9
5	10	16.1
6	7	11.3
Total	59	95.2
Systems	3	4.8
Total	62	100.0

Appendix 128. Perceived Difficulty of Mobile Phone Number for KT

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	4	4.00	1.414	.707	1.75	6.25	3	6
Incorrect	41	3.76	1.578	.246	3.26	4.25	1	6
Not Found	9	1.44	1.014	.338	.67	2.22	1	4
Total	54	3.39	1.709	.233	2.92	3.86	1	6

Appendix

ANOVAs for KT

Appendix 129. KT Mother's Maiden Name ANOVA

KT Mother's Maiden Name ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	1.665	2	.832	.631	.536
	Within Groups	73.895	56	1.320		
	Total	75.559	58			
Maiden Time	Between Groups	2529.320	2	1264.660	3.502	.037
	Within Groups	20224.816	56	361.157		
	Total	22754.136	58			
Maiden Difficulty	Between Groups	62.731	2	31.366	46.000	.000
	Within Groups	38.184	56	.682		
	Total	100.915	58			

Appendix 130. KT Nickname ANOVA

KT Nickname ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	7.036	3	2.345	1.780	.162
	Within Groups	68.518	52	1.318		
	Total	75.554	55			
Nickname Time	Between Groups	148.714	3	49.571	.455	.715
	Within Groups	5667.840	52	108.997		
	Total	5816.554	55			
	Between Groups	78.650	3	26.217	21.233	.000

Nickname Difficulty	Within Groups	64.207	52	1.235		
	Total	142.857	55			

Appendix 131. KT Children's Names ANOVA

KT Children's Names ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	4.784	4	1.196	.879	.483
	Within Groups	72.113	53	1.361		
	Total	76.897	57			
Children Time	Between Groups	362.145	4	90.536	.567	.688
	Within Groups	8305.995	52	159.731		
	Total	8668.140	56			
Children Difficulty	Between Groups	83.370	4	20.843	15.779	.000
	Within Groups	70.009	53	1.321		
	Total	153.379	57			

Appendix 132. KT Pet's Names ANOVA

KT Pet's Names ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	3.858	2	1.929	1.259	.294
	Within Groups	68.955	45	1.532		
	Total	72.813	47			
Pet Time	Between Groups	185.194	2	92.597	.576	.567
	Within Groups	6595.056	41	160.855		

	Total	6780.250	43			
Pet Difficulty	Between Groups	46.813	2	23.406	18.160	.000
	Within Groups	58.000	45	1.289		
	Total	104.813	47			

Appendix 133. KT Middle Name ANOVA

KT Middle Name ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	3.611	3	1.204	.897	.449
	Within Groups	73.779	55	1.341		
	Total	77.390	58			
Middle Name Time	Between Groups	139.423	3	46.474	.336	.799
	Within Groups	7327.419	53	138.253		
	Total	7466.842	56			
Middle Name Difficulty	Between Groups	63.134	3	21.045	13.288	.000
	Within Groups	87.103	55	1.584		
	Total	150.237	58			

Appendix 134. KT Mobile Phone Number ANOVA

KT Mobile Phone ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	1.007	2	.503	.340	.713
	Within Groups	75.530	51	1.481		
	Total	76.537	53			

Mobile Phone Time	Between Groups	833.086	2	416.543	2.899	.064
	Within Groups	7183.631	50	143.673		
	Total	8016.717	52			
Mobile Phone Difficulty	Between Groups	41.050	2	20.525	9.200	.000
	Within Groups	113.783	51	2.231		
	Total	154.833	53			

Appendix 136. Accuracy of Mother's Maiden Name for OV

	Frequency	Percent
Missing	2	3.8
Correct	4	7.5
Partially Correct/Incomplete	2	3.8
Incorrect	16	30.2
Not found	29	54.7
Total	53	100.0

Appendix 137. Familiarity of Mother's Maiden Name Compared by Accuracy for OV

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	4	2.75	1.708	.854	.03	5.47	1	5
Partially Correct/Incomplete	2	4.00	.000	.000	4.00	4.00	4	4
Incorrect	16	3.13	1.025	.256	2.58	3.67	1	5
Not Found	29	3.00	1.134	.211	2.57	3.43	1	5
Total	51	3.06	1.121	.157	2.74	3.37	1	5

Appendix 138. Time of Mother's Maiden Name Compared by Accuracy for OV

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	4	17.75	8.578	4.289	4.10	31.40	10	30
Partially Correct/Incomplete	2	15.00	.000	.000	15.00	15.00	15	15
Incorrect	15	32.93	32.664	8.434	14.84	51.02	5	120
Not Found	29	18.24	11.385	2.114	13.91	22.57	5	60
Total	50	22.48	20.775	2.938	16.58	28.38	5	120

Appendix 139. Locations of Mother's Maiden Name for OV

Location	Number
Ancestry	1
AOL	1
Bing	2
Birth Records	1
Blog	0
Facebook	17
Google	5
LinkedIn	4
Phone	0
Quanki	1
Spokeo	2
Snapchat	0
Twitter	0
University of Washington	4
White Pages	3
Wikipedia	1
Yellow Pages	1

Appendix 140. Perceived Difficulty of Mother's Maiden Name by Difficulty Score for OV

	Frequency	Percent
1	33	62.3
2	3	5.7
3	2	3.8
4	8	15.1
5	5	9.4
Total	51	96.2
Missing	2	3.8
Total	53	100.0

Appendix 141. Perceived Difficulty of Mother's Maiden Name by Correctness for OV

				95% Confidence Interval for Mean			
N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max

Correct	4	4.25	.500	.250	3.45	5.05	4	5
Partially Correct/Incomplete	2	4.50	.707	.500	-1.85	10.85	4	5
Incorrect	16	2.81	1.601	.400	1.96	3.67	1	5
Not Found	29	1.07	.371	.069	.93	1.21	1	3
Total	51	2.00	1.497	.210	1.58	2.42	1	5

Appendix 142. Accuracy of Nickname for OV

	Frequency	**Percent**
Missing	7	13.2
Correct	15	28.3
Incorrect	25	47.2
Not found	6	11.3
Total	53	100.0

Appendix 143. Familiarity Compared by Accuracy of Nickname for OV

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	15	2.53	1.060	.274	1.95	3.12	1	4
Incorrect	25	3.16	1.028	.206	2.74	3.58	1	5
Not found	6	3.00	.894	.365	2.06	3.94	2	4
Total	46	2.93	1.041	.154	2.63	3.24	1	5

Appendix 144. Time Compared by Accuracy of Nickname for OV

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	15	15.13	12.800	3.305	8.04	22.22	2	50
Incorrect	25	19.64	18.009	3.602	12.21	27.07	1	60
Not found	6	14.67	6.218	2.539	8.14	21.19	3	20
Total	46	17.52	15.288	2.254	12.98	22.06	1	60

Appendix 145. Locations of Nickname for OV

Location	**Number**
Ancestry	0
AOL	0
Bing	1

Birth Records	0
Blog	1
Facebook	28
Google	3
LinkedIn	0
Phone Book (unspecified)	0
Quanki	1
Spokeo	1
Snapchat	0
Twitter	3
University	0
White Pages	4
Wikipedia	1
Yellow Pages	0

Appendix 146. Perceived Difficulty of Nickname for OV

	Frequency	**Percent**
1	15	28.3
2	9	17.0
3	10	18.9
4	8	15.1
5	3	5.7
6	3	5.7
Total	48	90.6
Missing	5	9.4
Total	53	100.0

Appendix 147. Perceived Difficulty of Nickname by Accuracy for OV

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	15	2.47	1.457	.376	1.66	3.27	1	5
Incorrect	25	3.24	1.480	.296	2.63	3.85	1	6
Not found	6	1.33	.816	.333	.48	2.19	1	3
Total	46	2.74	1.527	.225	2.29	3.19	1	6

Appendix 148. Accuracy of Children's Names for OV

	Frequency	Percent
Missing	2	3.8
Correct	40	75.5
Incorrect	5	9.4
Not found	6	11.3
Total	53	100.0

Appendix 149. Accuracy of Children's Names by Accuracy Group for OV

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	40	3.05	1.061	.168	2.71	3.39	1	5
Incorrect	5	2.80	1.095	.490	1.44	4.16	1	4
Not Found	6	2.33	1.506	.615	.75	3.91	1	4
Total	51	2.94	1.121	.157	2.63	3.26	1	5

Appendix 150. Familiarity with Search of Children's Names by Accuracy Group for OV

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	40	3.05	1.061	.168	2.71	3.39	1	5
Incorrect	5	2.80	1.095	.490	1.44	4.16	1	4
Not Found	6	2.33	1.506	.615	.75	3.91	1	4
Total	51	2.94	1.121	.157	2.63	3.26	1	5

Appendix 151. Search time of Children's Names for OV

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max

Locations	Number
Ancestry	0
AOL	0
Bing	1
Birth Records	0
Blog	1
Facebook	28
Google	3
LinkedIn	0
Phone	0
Quanki	1
Spokeo	1
Snapchat	0
Twitter	3
University of Washington	0
Wikitree	1
White Pages	4
Wikipedia	1

Correct	40	14.08	12.190	1.927	10.18	17.97	1	60
Incorrect	5	11.00	10.840	4.848	-2.46	24.46	5	30
Not Found	6	18.50	9.670	3.948	8.35	28.65	10	30
Total	51	14.29	11.743	1.644	10.99	17.60	1	60

Appendix 152. Locations of Children's Names for OV

Appendix 153. Perceived Difficulty of Children's Names' for OV

	Frequency	Percent
1	13	24.5
2	8	15.1
3	7	13.2
4	13	24.5
5	9	17.0
6	2	3.8
Total	52	98.1
Missing	1	1.9
Total	53	100.0

Appendix 154. Perceived Difficulty of Children's Names by Accuracy for OV

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	40	3.28	1.450	.229	2.81	3.74	1	6
Incorrect	5	4.20	1.304	.583	2.58	5.82	3	6
Not Found	6	1.00	.000	.000	1.00	1.00	1	1
Total	51	3.10	1.565	.219	2.66	3.54	1	6

Appendix 155. Accuracy of Pet's Names for OV

	Frequency	Percent
Missing	8	15.1
Correct	35	66.0
Not found	10	18.9
Total	53	100.0

Appendix 156. Familiarity with search of Pet's Names for OV

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	35	3.06	1.162	.196	2.66	3.46	1	5
Not found	10	2.70	1.059	.335	1.94	3.46	1	4
Total	45	2.98	1.138	.170	2.64	3.32	1	5

Appendix 157. Time of Pet's Names for OV

	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Correct	33	13.97	12.231	2.129	9.63	18.31	2	60
Not Found	10	9.20	4.803	1.519	5.76	12.64	2	15
Total	43	12.86	11.094	1.692	9.45	16.27	2	60

Appendix 158. Locations of Pet's Names for OV

Location	**Number**
Ancestry	0

AOL	0
Bing	1
Birth Records	0
Blog	2
Facebook	26
Google	2
LinkedIn	2
Phone	0
Quanki	0
Spokeo	1
Snapchat	1
Twitter	3
University of Washington	0
White Pages	0
Wikipedia	0
Yellow Pages	0

Appendix 159. Difficulty rating of Pet's Names for OV

	Frequency	**Percent**
1	17	32.1
2	9	17.0
3	6	11.3
4	12	22.6
5	2	3.8
Total	46	86.8
Missing	7	13.2
Total	53	100.0

Appendix 160. Perceived Difficulty of Pet's Names by Accuracy Group for OV

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	35	2.86	1.240	.210	2.43	3.28	1	5
Not Found	10	1.00	.000	.000	1.00	1.00	1	1
Total	45	2.44	1.341	.200	2.04	2.85	1	5

Appendix 161. Accuracy of Middle Name for OV

	Frequency	**Percent**

Missing	2	3.8
Correct	24	45.3
Correct/Incomplete	15	28.3
Incorrect	9	17.0
Not found	3	5.7
Total	53	100.0

Appendix 162. Familiarity by Accuracy of Middle Name for OV

					95% Confidence Interval for Mean		
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	24	2.92	1.139	.232	2.44	3.40	1	5
Correct/Incomplete	15	2.93	1.033	.267	2.36	3.51	1	4
Incorrect	9	2.89	1.054	.351	2.08	3.70	1	4
Not Found	3	3.33	2.082	1.202	-1.84	8.50	1	5
Total	51	2.94	1.121	.157	2.63	3.26	1	5

Appendix 163. Time by Accuracy of Middle Name for OV

					95% Confidence Interval for Mean		
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	24	12.08	10.886	2.222	7.49	16.68	2	45
Correct/Incomplete	15	11.53	10.842	2.799	5.53	17.54	1	35
Incorrect	9	20.44	15.645	5.215	8.42	32.47	3	56
Not Found	3	16.33	12.342	7.126	-14.33	46.99	6	30
Total	51	13.65	11.998	1.680	10.27	17.02	1	56

Appendix 164. Locations of Middle Name for OV

Number	**Location**
Academia	1
Ancestry	0
AOL	0
Bing	0
Been Verified	1

Birth Records	0
Blog	0
Facebook	17
Google	2
LinkedIn	2
Phone	0
Quanki	1
Research Gate	1
Spokeo	1
Snapchat	0
Twitter	6
University of Washington	2
White Pages	19
Wikipedia	0
Yellow Pages	0

Appendix 165. Perceived Difficulty of Middle Name for OV

	Frequency	**Percent**
1	8	15.1
2	9	17.0
3	13	24.5
4	5	9.4
5	9	17.0
6	7	13.2
Total	51	96.2
System	2	3.8
Total	51	96.2

Appendix 166. Perceived Difficulty by Accuracy of Middle Name for OV

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	24	3.71	1.398	.285	3.12	4.30	1	6
Correct/Incomplete	15	3.93	1.580	.408	3.06	4.81	2	6
Incorrect	9	2.33	1.732	.577	1.00	3.66	1	6
Not Found	3	1.00	.000	.000	1.00	1.00	1	1
Total	51	3.37	1.661	.233	2.91	3.84	1	6

Appendix 167. Accuracy of Mobile Phone Number for OV

	Frequency	Percent
Missing	9	17.0
Correct	5	9.4
Correct/Incomplete	1	1.9
Incorrect	7	13.2
Not found	31	58.5

Appendix 168. Familiarity by Accuracy of Mobile Phone Number for OV

					95% Confidence Interval for Mean		
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	5	3.00	1.000	.447	1.76	4.24	2	4
Correct/Incomplete	1	3.00	.	.	.	.	3	3
Incorrect	7	2.43	.976	.369	1.53	3.33	1	3
Not Found	31	3.23	1.117	.201	2.82	3.64	1	5
Total	44	3.07	1.087	.164	2.74	3.40	1	5

Appendix 169. Mobile Phone Number Search Time by Accuracy Group

					95% Confidence Interval for Mean		
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	5	25.00	36.401	16.279	-20.20	70.20	5	90
Correct/Incomplete	1	5.00	.	.	.	.	5	5
Incorrect	7	30.14	16.036	6.061	15.31	44.97	13	60
Not Found	31	19.90	21.212	3.810	12.12	27.68	1	120
Total	44	21.77	22.243	3.353	15.01	28.54	1	120

Appendix 170. Locations of *Mobile Phone Number* for OV

Location	Number
Ancestry	0
AOL	0
Bing	1
Birth Records	0
Blog	0
Facebook	6
Google	2
LinkedIn	1

Nuwber	1
Phone	1
Pipl	1
Quanki	0
Spokeo	1
Snapchat	0
Truthfinder	1
Twitter	1
University of Washington	1
White Pages	3
Wikipedia	0
Yellow Pages	1

Appendix 171. Frequency of Perceived Difficulty of Mobile Phone Number for OV

	Frequency	**Percent**
1	31	58.5
2	4	7.5
3	5	9.4
4	3	5.7
5	2	3.8
Total	45	84.9
Missing	8	15.1
Total	53	100.0

Appendix 172. Perceived Difficulty of Mobile Phone Number for OV

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Correct	5	3.00	1.225	.548	1.48	4.52	1	4
Correct/Incomplete	1	5.00	.	.	.	.	5	5
Incorrect	7	2.57	1.512	.571	1.17	3.97	1	5
Not Found	31	1.19	.543	.097	.99	1.39	1	3
Total	44	1.70	1.193	.180	1.34	2.07	1	5

ANOVAs for OV

Appendix 173. OV Mother's Maiden Name ANOVA

OV Mother's Maiden Name ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	2.324	3	.775	.602	.617
	Within Groups	60.500	47	1.287		
	Total	62.824	50			
Maiden Time	Between Groups	2361.486	3	787.162	1.927	.138
	Within Groups	18786.994	46	408.413		
	Total	21148.480	49			
Maiden Difficulty	Between Groups	68.450	3	22.817	24.625	.000
	Within Groups	43.550	47	.927		
	Total	112.000	50			

Appendix 174. OV Nickname ANOVA

OV Nickname ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	3.711	2	1.856	1.769	.183
	Within Groups	45.093	43	1.049		
	Total	48.804	45			
Nickname Time	Between Groups	246.652	2	123.326	.516	.600
	Within Groups	10270.827	43	238.856		
	Total	10517.478	45			
Nickname Difficulty	Between Groups	19.243	2	9.621	4.832	.013
	Within Groups	85.627	43	1.991		

Total		104.870	45			

Appendix 175. OV Children's Names ANOVA

OV Children's Names ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	2.790	2	1.395	1.115	.336
	Within Groups	60.033	48	1.251		
	Total	62.824	50			
Children Time	Between Groups	162.313	2	81.157	.579	.565
	Within Groups	6732.275	48	140.256		
	Total	6894.588	50			
Children Difficulty	Between Groups	33.735	2	16.867	9.120	.000
	Within Groups	88.775	48	1.849		
	Total	122.510	50			

Appendix 176. OV Pet's Names ANOVA

OV Pet's Names ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	.992	1	.992	.762	.388
	Within Groups	55.986	43	1.302		
	Total	56.978	44			
Pet's Time	Between Groups	174.593	1	174.593	1.433	.238
	Within Groups	4994.570	41	121.819		
	Total	5169.163	42			
Pet's Difficulty	Between Groups	26.825	1	26.825	22.061	.000
	Within Groups	52.286	43	1.216		

Total		79.111	44			

Appendix 177. OV Middle Name ANOVA

OV Middle Name ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	.501	3	.167	.126	.944
	Within Groups	62.322	47	1.326		
	Total	62.824	50			
Middle Name Time	Between Groups	563.192	3	187.731	1.330	.276
	Within Groups	6634.456	47	141.159		
	Total	7197.647	50			
Middle Name Difficulty	Between Groups	34.030	3	11.343	5.132	.004
	Within Groups	103.892	47	2.210		
	Total	137.922	50			

OV Mobile Phone ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Familiarity	Between Groups	3.662	3	1.221	1.036	.387
	Within Groups	47.134	40	1.178		
	Total	50.795	43			
Mobile Phone Time	Between Groups	932.160	3	310.720	.611	.612
	Within Groups	20341.567	40	508.539		
	Total	21273.727	43			
Mobile Phone Difficulty	Between Groups	32.606	3	10.869	15.226	.000
	Within Groups	28.553	40	.714		

Total	61.159	43			

Bibliography

Ackerman, M. S. (2000). The intellectual challenge of CSCW: The gap between social requirements and technical feasibility. *Human-Computer Interaction, 15*(2-3), 179-204.

Ackerman, M. S. (2004). Privacy in pervasive environments: next generation labeling protocols. *Personal & Ubiquitous Computing, 8*(6), 430-439.

Acquisti, A., Adjerid, I., & Brandimarte, L. (2013). Gone in 15 seconds: The limits of privacy transparency and control. *IEEE Security & Privacy, 11*(4), 72-74.

Acquisti, A., Friedman, A., & Telang, R. (2006). Is there a cost to privacy breaches? An event study. *ICIS 2006 Proceedings*, 94.

Acquisti, A., & Gross, R. (2006). *Imagined communities: Awareness, information sharing and privacy on the facebook*. Berlin Heidelberg: Springer.

Adamic, L. A., & Adar, E. (2003). Friends and neigGCors on the web. *Social Networks, 25*(3), 211-230.

Adams, A., & Sasse, M. (1999). Users are not the enemy. *Communications of the ACM, 49*(12), 41-46.

Adams, C., & Dimitrinu, A. (2008). A two-phase authentication protocol using the cell phone as a token. *Journal Of Information Privacy & Security, 4*(2), 23-39.

Alazab, A., Abawajy, J., Hobbs, M., & Khraisat, A. (2013). Crime toolkits: The current threats to web applications. *Journal Of Information Privacy & Security, 9*(2), 21-39.

Amar, R., & Stasko, J. (2004). *A knowledge task-based framework for design and evaluation of information visualizations.* Paper presented at the IEEE Symposium on Information Visualization, 2004.

Anderson, C., & Agarwal, R. (2010). Practicing safe computing: A multimethod empirical examination of home computer user security behavioral intentions. *MIS Quarterly, 34*(3), 613-643.

Andrews, L. W. (2002). Passwords reveal your personality. *Psychology Today, 35*(16).

Armando, A., Carbone, R., Compagna, L., Cuéllar, J., Pellegrino, G., & Sorniotti, A. (2013). An authentication flaw in browser-based single sign-on protocols: Impact and remediations. *Computers & Security, 33*, 41-58.

Baddeley, A. (1994). The magical number seven: Still magic after all these years? *Psychological Review, 101*(2), 353-356.

Bardram, J. E. (2005). The trouble with login: On usability and computer security in ubiquitous computing. *Personal and Ubiquitous Computing, 9*(6), 357-367.

Belanger, F. F., & Crossier, R. E. (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly, 35*(4), 1017-1041.

Beldad, A., de Jong, M., & Steehouder, M. (2011). A comprehensive theoretical framework for personal information-related behaviors on the internet. *Information Society, 27*(4), 220-232.

Benson, V., Saridakis, G., & Tennakoon, H. (2015). Information disclosure of social media users: Does control over personal information, user awareness and security notices matter? *Information Technology & People, 28*(3), 426-441.

Berendt, B., Günther, O., & Spiekermann, S. (2005). Privacy in e-commerce: Stated preferences vs. actual behavior. *Communications of the ACM, 48*(4), 101-106.

Beresford, A. R., Kübler, D., & Preibusch, S. (2012). Unwillingness to pay for privacy: A field experiment. *Economics Letters, 117*(1), 25-27.

Besnard, D., & Arief, B. (2004). Computer security impaired by legitimate users. *Computers & Security, 23*, 253-264.

Biddle, R., Chiasson, S., & Van Orschot, P. C. (2012). Graphical passwords: Learning from the first twelve years. *ACM Computing Surveys, 44*(4), 19:11-19:41.

Bonneau, J., Herley, C., van Oorschot, P. C., & Stajano, F. (2012). The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. *IEEE Symposium on Security and Privacy*, 553-567.

Bonneau, J., Herley, C., Van Oorschot, P. C., & Stajano, F. (2015). Passwords and the evolution of imperfect authentication. *Communications of the ACM, 58*(7), 78-87.

Bonneau, J., Just, M., & Matthews, G. (2010). What's in a name? Evaluating statistical attacks on personal knowledge questions. In D. E. Lewis (Ed.), *Financial Cryptography and Data Security* (pp. 98-113). Berlin: Springer.

Borchert, C. J., Pinguelo, F. M., & Thaw, D. (2014). Resasonable expecations of privacy settings: Social media and the stored communications act. *Duke Law & Technology Review, 13*(1), 36-65.

Brainard, J., Juels, A., Rivest, R. L., Szydlo, M., & Yung, M. (2006). *Fourth-factor authentication: Somebody you know.* Paper presented at the ACM Conference on Computer and Communications Security.

Brenner, J., & Smith, A. (2013). 72% of online adults are social networking site users. *Washington, DC: Pew Internet & American Life Project*.

Brown, A. S., Bracken, E., Zoccoli, S., & Douglas, K. (2004). Generating and remembering passwords. *Applied Cognitive Psychology, 18*(6), 641-651.

Bunn, A. (2015). The curious case of the right to be forgotten. *Computer Law & Security Review: The International Journal of Technology Law and Practice, 31*, 336-350.

Caldwell, T. (2014). Feature: The true cost of being hacked. *Computer Fraud & Security, 2014*, 8-13.

Calo, M. l. (2011). The boundaries of privacy harm. *Indiana Law Journal, 86(3)*, 1131-1162.

Campbell, J., Ma, W., & Kleeman, D. (2011). Impact of restrictive composition policy on user password choices. *Behaviour & Information Technology, 30*(3), 379-388.

Campbell, K., Gordon, L. A., Loeb, M. P., & Zhou, L. (2003). The economic cost of publicly announced information security breaches: Empirical evidence from the stock market. *Journal of Computer Security, 11*(3), 431-448.

Chiasson, S., Stobert, E., Forget, A., Biddle, R., & Van Oorschot, P. C. (2012). Persuasive cued click-points: Design, implementation, and evaluation of a knowledge-based authentication mechanism. *IEEE Transactions on Dependable and Secure Computing, 9*(2), 222-235.

Cockcroft, S., & Heales, J. (2005). National culture, trust and internet privacy concerns. *ACIS 2005 Proceedings*, 65.

Connelly, C., Archer, N., Yuan, Y., & Guo, K. (2011). Understanding nonmalicious security violations in the workplace: A composite behavior model. *Journal of Management Information Systems, 28*(2), 203-236.

Connelly, D. P., Rich, E. C., Curley, S. P., & Kelly, J. T. (1990). Knowledge resource preferences of family physicians. *The Journal Of Family Practice, 30*(3), 353-359.

Corbett, S. (2013). The retention of personal information online: A call for international regulation of privacy law. *Computer Law and Security Review: The International Journal of Technology and Practice, 29*, 246-254. dCowan, N. (2015). George miller's magical number of immediate memory in retrospect: Observations on the faltering progression of science. *Psychological Review*.

Custers, B., van der Hof, S., & Schermer, B. (2014). Privacy expectations of social media users: The role of informed consent in privacy policies. *Policy & Internet, 6*(3), 268-295.

Dhamija, R., & Perrig, A. (2000). *Deja Vu-A User Study: Using Images for Authentication.* Paper presented at the USENIX Security Symposium.

Dillon, A. (1992). Reading from paper versus screens: A critical review of the empirical literature. *Ergonomics, 35*(10), 1297-1326.

Dlamini, M. T., Eloff, J. P., & Eloff, M. M. (2009). Information security: The moving target. *Computers & Security, 28*(3/4), 189-198.

Drahansky, M., Brezinova, E., Hejtmankova, D., & Orsag, F. (2010). Fingerprint recognition influenced by skin diseases. *International Journal of Bio-Science & Bio-Technology, 2*(4), 11-21.

Duggan, G. B., Johnson, H., & Grawemeyer, B. (2012). Rational security: Modelling everyday password use. *International Journal of Human-Computer Studies, 70*(6), 415-431.

Faiola, A. (2007). The design enterprise: Rethinking the HCI education paradigm. *Design Issues, 23*(3), 30-45.

Faundez-Zanuy, M. (2004). On the vulnerability of biometric security systems. *Aerospace and Electronic Systems Magazine, IEEE, 19*(6), 3-8.

Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers, 22*(5), 323-327.

Furnell, S. (2005). Authenticating ourselves: Will we ever escape the password? *Network Security, 2005*(3), 8-13.

Furnell, S. (2007). A comparison of website user authentication mechanisms. *Computer Fraud & Security, 2007*(9), 5-9.

Furnell, S., & Zekri, L. (2006). Replacing passwords: In search of the secret remedy. *Network Security, 2006*(1), 4-8.

Gao, H., Jia, W., Ye, F., & Ma, L. (2013). A survey on the use of graphical passwords in security. *Journal of Software, 8*(7), 1678-1698.

Gerlach, J., Widjaja, T., & Buxmann, P. (2015). Handle with care: How online social network providers' privacy policies impact users' information sharing behavior. *Journal of Strategic Information Systems, 24*, 33-43.

Gignac, G. E. (2015). The magical numbers 7 and 4 are resistant to the Flynn effect: No evidence for increases in forward or backward recall across 85 years of data. *Intelligence, 48*, 85-95.

Gong, N., & Wang, D. (2014). On the security of trusee-based social authentication. *IEEE Transactions on Information Forensics and Security, 9*(8), 1251-1263.

Grawemeyer, B., & Johnson, H. (2011). Using and managing multiple passwords: A week to a view. *Interacting with Computers, 23*(3), 256-267.

Griffith, V., & Jakobsson, M. (2005). Messin' with texas deriving mother's maiden names using public records. In J. Ioannidis, A. Keromytis, & M. Yung (Eds.), *Applied Cryptography and Network Security: Third International Conference, ACNS 2005, New York, NY, USA, June 7-10, 2005. Proceedings* (pp. 91-103). Berlin: Springer.

Gross, R., & Acquisti, A. (2005). Information revelation and privacy in online social networks. *In Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, 71-80.

Grossman, W. (2009). Strategy: The user is not the enemy: How to increase infosecurity usability. *Infosecurity, 6*(5), 20-22.

Grudin, J. (2008). A moving target: The evolution of HCI. In A. J. Sears, J. (Ed.), *The human computer interaction handbook: Fundamentals, evolving technologies, and emerging applications, 2nd Ed.* (pp. 1-24). New York: Lawrence Erlbaum Associates.

Gulenko, I. (2014). Improving passwords: Influence of emotions on security behavior. *Information Management & Computer Security, 22*(2), 167-178.

Gupta, M., & Sharman, R. (2012). Determinants of data breaches: A categorization-based empirical investigation. *Journal Of Applied Security Research, 7*(3), 375-395.

Hall, J. L., & McGraw, D. (2014). For telehealth to succeed, privacy and security risks must be identified and addressed. *Health Affairs, 33*(2), 216-221.

Hartzog, W., & Stutzman, F. (2013). Obscurity by design. *Washington Law Review, 88*(2), 386-418.

Heckle, R. R., & Lutters, W. G. (2011). Tensions of network security and collaborative work practice: Understanding a single sign-on deployment in a regional hospital. *Internation Journal of Medical Informatics, 80*(8), e49-61.

Herath, T., & Rao, H. R. (2009). Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. *Decision Support Systems, 47*(2), 154-165.

Herley, C., & Van Oorschot, P. (2012). A research agenda acknowledging the persistence of passwords. *IEEE Security & Privacy, 10*(1), 28-36.

Hoehle, H., & Venkatesh, V. (2015). Mobile application usability: Conceptualization and instrument development. *MIS Quarterly, 39*(2), 435-A412.

Il-Horn, H., Kai-Lung, H. U. I., Sang-Yong Tom, L. E. E., & Png, I. P. L. (2007). Overcoming online information privacy concerns: An information-processing theory approach. *Journal of Management Information Systems, 24*(2), 13-42.

Jain, A., Hong, L., & Pankanti, S. (2000). Biometric identification. *Communications of the ACM, 43*(2), 90-98.

Jones, W. (2004). Finders, keepers? The present and future perfect in support of personal information management. *First Monday, 9*(3).

Jourard, S. M. (1966). Some Psychological Aspects of Privacy, 307.

Kaikkonen, A., Kekalainen, A. , Cankar, M., . . . Kankainen, A. (2005). Usability testing of mobile applications: a comparison between laboratory and field testing. *J. Usability Studies, 1*(1), 4-16.

Keith, M., Shao, B., & Steinbart, P. J. (2007). The usability of passphrases for authentication: An empirical field study. *International Journal of Human-Computer Studies, 65*(1), 17-28.

Kim, K.-S., Sin, S.-C. J., & Tsai, T.-I. (2014). Individual Differences in Social Media Use for Information Seeking. *The Journal of Academic Librarianship, 40*, 171-178.

Kline, D. M., He, L., & Yaylacicegi, U. (2011). User perceptions of security technologies. *International Journal of Information Security and Privacy, 5*(2), 1-12.

Kokolakis, S. (2015). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*.

Leach, J. (2003). Improving user security behaviour. *Computers & Security, 22*(8).

Li, J. (2014). Data protection in healthcare social networks. *IEEE Software, 31*(1), 46-53.

Lindamood, J., Heatherly, R., Kantarcioglu, M., & Thuraisingham, B. (2009). *Inferring private information using social network data*. Paper presented at the Proceedings of the 18th international conference on world wide web, Madrid, Spain.

Liu, Y., Gummadi, K. P., Krishnamurthy, B., & Mislove, A. (2011). *Analyzing facebook privacy settings: User expectations vs. reality*. Paper presented at the Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference, Berlin, Germany.

Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 4*(2), 73-79.

Madejski, M., Johnson, M., & Bellovin, S. M. (2012). A study of privacy settings errors in an online social network. *2012 IEEE International Conference on Pervasive Computing & Communications Workshops*, 340.

Medlin, B. D. (2013). Social engineering techniques and password security: two issues relevant in the case of health care workers. *International Journal of Cyber Warfare and Terrorism, 3*(2), 58-70.

Medlin, B. D., & Cazier, J. A. (2005). An investigative study: Consumers password choices on an e-commerce site. *Journal Of Information Privacy & Security, 1*(4), 33-52.

Miller, G. (1956). The magical number seven, plus or minus two some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Morris, R., Thompson, K., & Gaines, R. S. (1979). Password Security: A Case History. *Communications of the ACM, 22*(11), 594-597.

Muhammad, L. J., Garba Ali, A., & Iliya, I. S. (2015). Security challenges for building knowledge-based economy in nigeria. *International Journal of Security & Its Applications, 9*(1), 119-124.

Ng, B., Kankanhalli, A., & Xu, Y. (2009). Studying users' computer security behavior: A health belief perspective. *Decision Support Systems, 46*(4), 815-825.

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs, 41*(1), 100-126.

Noyes, J. M., & Garland, K. J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics, 51*(9), 1352-1375.

O'Gorman, L. (2003). Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE, 91*(12), 2021-2040.

Oravec, J. A. (2012). Deconstructing "personal privacy" in an age of social media: Information control and reputation mangement dimensions. *International Journal of the Academic Business World, 6*(1), 95-104.

Ostertagova, E., Ostertag, O., & Kova´c, J. (2014). Methodology and application of the kruskal-wallis test. *Applied Mechanics and Materials, 611*, 115-120.

Palmer, V. V. (2011). Three milestones in the history of privacy in the united states. *Tulane European & Civil Law Forum, 26*(1), 67-97.

Park, D., Boyd, C., & Dawson, E. (2000). *Classification of authentication protocols: A practical approach.* Paper presented at the Information Security. Third International Workshop, ISW 2000., Berlin.

Pavlou, P. A. (2011). State of the information privacy literature: Where are we now and where should we go? *MIS Quarterly, 35*(4), 977-988.

Petronio, S. (2002). *Boundaries of privacy: Dialectics of disclosure*. Albany, NY: State University of New York Press.

Piccolotto, P., & Patricio, M. (2014). Biometrics from the user point of view: Deriving design principles from user perceptions and concerns about biometric systems. *Intel Technology Journal, 18*(4), 30-44.

Polakis, I., Ilia, P., Maggi, F., Lancini, M., Kontaxis, G., Zanero, S., . . . Keromytis, A. D. (2014). *Faces in the distorting mirror: Revisiting photo-based social authentication*. Paper presented at the Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, Arizona, USA.

Polakis, I., Lancini, M., Kontaxis, G., Maggi, F., Ioannidis, S., Keromytis, A. D., & Zanero, S. (2012). *All your face are belong to us: breaking Facebook's social authentication*. Paper presented at the Proceedings of the 28th Annual Computer Security Applications Conference, Orlando, Florida, USA.

Powell, C. D. (2011). "You already have zero privacy. get over it!" Would warren and brandeis argue for privacy for social networking? *Pace Law Review, 31*(1), 146-181.

Preibusch, S. (2013). Guide to measuring privacy concern: Review of survey and observational instruments. *International Journal of Human-Computer Studies, 71*(12), 1133-1143.

Qinghan, X. (2005). *Security issues in biometric authentication.* Paper presented at the Proceedings from the Sixth Annual IEEE Systems, Man and Cybernetics (SMC) Information Assurance Workshop., Piscataway, NJ, USA.

Qiong, X. (2013). Diffusion of social media adoption in everyday academic information seeking. *International Journal of Technology, Knowledge & Society*, *9*(4), 41-60.

Rabkin, A. (2008). Personal knowledge questions for fallback authentication. *ACM International Conference Proceeding Series*, 13.

Rada, R. (2008). *Information systems and healthcare enterprises.* Hersey, NY: IGI Publishing.

Reeder, R., & Schechter, S. (2011). When the password doesn't work: Secondary authentication for websites. *IEEE Security & Privacy Magazine, 9*(2), 43.

Rhee, H., Kim, C., & Ryu, Y. U. (2009). Self-efficacy in information security: Its influence on end users' information security practice behavior. *Computers & Security, 28*(8), 816-826.

San-Tsai, S., Pospisil, E., Muslukhov, I., Dindar, N., Hawkey, K., & Beznosov, K. (2013). Investigating users' perspectives of web single sign-on: Conceptual gaps and acceptance model. *ACM Transactions on Internet Technology, 13*(1), 1-35.

Sasse, M., Brostoff, S., & Weirich, D. (2001). Transforming the 'weakest link' a human-computer interaction approach to usable and effective security. *BT Technology Journal, 19*(3).

Sauro, J., & Lewis, J. R. (2010). *Average task times in usability tests: What to report?*

Schau, H. J., & Gilly, M. C. (2003). We are what we post? Self-presentation in personal web space. *Journal of Consumer Research, 30*(3), 385-404.

Schechter, S., Brush, A. J. B., & Egelman, S. (2009). Its no secret: Measuring the reliability of authentication via 'secret' questions. *Proceedings of the 2009 30th IEEE Symposium on Security and Privacy*, 375-390.

Schechter, S., Egelman, S., & Reeder, S. (2009). *It's not what you know, but who you know: a social approach to last-resort authentication*. Paper presented at the

Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA.

Schneier, B. (2010). Schneier on security: Privacy and control. *Journal of Privacy and Confidentiality, 2*(1), 3-4.

Schouten, A. P., Valkenburg, P. M., & Peter, J. (2009). An Experimental Test of Processes Underlying Self-Disclosure in Computer-Mediated Communication. *Cyberpsychology, 3*(2), 1-13.

Sharma, S., & Crossler, R. E. (2014). Disclosing too much? Situational factors affecting information disclosure in social commerce environment. *Electronic Commerce Research and Applications, 13*, 305-319.

Shay, R., Komanduri, S., Kelley, P. G., Leon, P. G., Mazurek, M. L., Bauer, L., . . . Cranor, L. F. (2010). *Encountering stronger password requirements: user attitudes and behaviors*. Paper presented at the Proceedings of the Sixth Symposium on Usable Privacy and Security, Redmond, Washington, USA.

Smith, H., Milberg, S. J., & Burke, S. J. (1996). Information privacy: Measuring individuals' concerns about organizational practices. *MIS Quarterly, 20*(2), 167-196.

Social Security Administration. (n.d.). Popularity of a name. Retrieved from https://www.ssa.gov/oact/babynames/index.html

Soliman, W., & Tuunainen, V. K. (2015). Understanding Continued Use of Crowdsourcing Systems: An Interpretive Study. *Journal of Theoretical & Applied Electronic Commerce Research, 10*(1), 1-18.

Stutzman, F. (2006). An evaluation of identity-sharing behavior in social network communities. *Journal of the International Digital Media and Arts Association, 3*(1), 10-18.

Sutanto, J., Palme, E., Chuan-Hoo, T., & Chee Wei, P. (2013). Addressing the personalization-privacy paradox: An empirical assessment from a field experiment on smartphone users. *MIS Quarterly, 37*(4), 1141-A1145.

Taiabul Haque, S. M., Wright, M., & Scielzo, S. (2014). Hierarchy of users′ web passwords: Perceptions, practices and susceptibilities. *International Journal of Human - Computer Studies, 72*, 860-874.

Tam, L., Glassman, M., & Vandenwauver, M. (2010). The psychology of password management: A tradeoff between security and convenience. *Behaviour & Information Technology, 29*(3), 233-244.

Toe, C. A. (2013). *An Examination of the Explicit Costs of Sensitive Information Security Breaches.* ProQuest LLC. Available from EBSCOhost eric database.

Toomim, M., Zhang, X., Fogarty, J., & Landay, J. A. (2008). Access control by testing for shared knowledge. *Conference on Human Factors in Computing Systems Proceedings*, 193.

Trewin, S., Swart, C., Koved, L., Martino, J., Singh, K., & Ben-David, S. (2012). *Biometric authentication on a mobile device: A study of user effort, error and task disruption*. Paper presented at the Proceedings of the 28th Annual Computer Security Applications Conference, Orlando, Florida, USA.

United States Census Bureau. (2010). *Surnames Occurring 100 or more times*. Retrieved from https://www.census.gov/topics/population/genealogy/data/2010_surnames.html

Vagias, W. M. (2006). Likert-type scale response anchors. *Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University.*

van Dijck, J. (2013). 'You have one identity': Performing the self on Facebook and LinkedIn. *Media, Culture & Society, 35*(2), 199-215.

Vander Veen, C. (2013). Why do we still use passwords? *Government Technology, 26*(11), 14-34.

Vickery, J. R. (2015). 'I don't have anything to hide, but … ': The challenges and negotiations of social and mobile media privacy for non-dominant youth. *Information, Communication & Society, 18*(3), 281-294.

Walrave, M., Utz, S., Schouten, A. P., & Heirman, W. (2016). Editorial: The state of online self-disclosure in an era of commodified privacy. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 10*(1).

Waters, S., & Ackerman, J. (2011). Exploring privacy management on Facebook: Motivations and perceived consequences of voluntary disclosure. *Journal of Computer-Mediated Communication, 17*(1), 101-115.

Weirich, D., & Sasse, M. A. (2001). *Pretty good persuasion: A first step towards effective password security in the real world.* Paper presented at the Proceedings of the 2001 workshop on New security paradigms.

Wilson, C. (2013). *Credible checklists and quality questionnaires : a User-centered design method*: Burlington : Elsevier Science, 2013.

Yardi, S., Feamster, N., & Bruckman, A. (2008). Photo-based authentication using social networks. *AMC*.

Youn, S. (2005). Teenagers' perceptions of online privacy and coping behaviors: a risk–benefit appraisal approach. *Journal of Broadcasting & Electronic Media*, *49*(1), 86-110.

Zeng, D., & Lusch, R. (2013). Big data analytics: Perspective shifting from transactions to ecosystems. *IEEE Intelligent Systems*, *28*(2), 2-5.