USER DATA DISCLOSURE BEHAVIOR ON SMART HOME DEVICES: UNIFYING THE PRIVACY PARADOX & THE PRIVACY CALCULUS MODEL

A Dissertation by

Kirsten Carter

Master of Arts, Wichita State University, 2017

Bachelor of Science, Indiana University-Bloomington, 2014

Submitted to the Department of Psychology
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

May 2021

USER DATA DISCLOSURE BEHAVIOR ON SMART HOME DEVICES:
UNIFYING THE PRIVACY PARADOX & THE PRIVACY CALCULUS MODEL

The following faculty members have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirement for the degree of Doctor of Philosophy with a major in Psychology.

_____
Lisa Vangsness, Committee Chair

_____
Carryl Baldwin, Committee Member

_____
Joel Suss, Committee Member

_____
Traci Hart, Committee Member

_____
Sergio Salinas, Committee Member

Accepted for the College of Liberal Arts & Sciences

_____
Andrew Hippisley, Dean

Accepted for the Graduate School

_____
Coleen Pugh, Dean

DEDICATION

To my dear husband, Jake; my parents, Derek and Shellie; my siblings Jessica, Charles, and William; my grandparents, Paul and Portia Sallee; and my beloved friends

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Lisa Vangsness, for welcoming me into her lab in my time of need and unconditionally supporting my research and career goals. Her enthusiasm and mentorship are what got me to where I am today. She is a true asset to the Human Factors discipline and I am honored to be her student. Her intelligence, kindness, and generosity go unmatched. I would also like to extend my gratitude to members of my committee, Traci Hart, Joel Suss, Sergio Salinas, and Carryl Baldwin for their mentorship and support on this project. I want to thank my lab mates Will Choi, Kevin Morales, and Jasmine Granados for making our lab… memorable. Last, but certainly not least, special thanks to Tiffany Leverenz, Inga Sogaard, Kyle Rexer, and Neethan Siva for being the professional (and personal) support system I needed throughout my time as a graduate student.

ABSTRACT

This research tested the Privacy Paradox for smart home devices. Ninety-one participants completed a survey to record their self-reported (i.e., anticipated) data disclosure and a behavioral task to observe their actual data disclosure behaviors. In the self-report task, participants were presented a series of disclosure scenarios and indicated whether they would disclose their personal information in that context. In the behavioral task, participants interacted with a functional prototype by completing a series of tasks typical of a smart home. The Privacy Calculus model was tested as an explanatory model of the Privacy Paradox within the context of IoT devices. Use case value, perceived risk of disclosure, privacy concerns, and trust in the smart home device were hypothesized predictors of self-reported and behavioral data disclosure.

The observed weak relationship between self-reported and behavioral disclosure provided evidence for the Privacy Paradox. Use case value, privacy concerns, and trust reliably predicted self-reported disclosure, while use case value and perceived risk of disclosure reliably predicted disclosure behaviors. Findings show that when it comes to smart homes, people’s disclosure behaviors do not match their self-reported disclosure and that use case value, perceived risk, privacy concerns, and trust serve as predictors to the Privacy Paradox.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

Today's technology is becoming increasingly connected to the internet and other devices, a concept of innovation known as the Internet of Things (IoT). The IoT integrates the physical and virtual world through the use of embedded hardware sensors and the Internet. Essentially, anyone can use the Internet to connect to any device from anywhere (Khan et al., 2012). This is made possible through the continual gathering, sending, and receiving of information from both the device's sensors and the user themselves (Atzori et al., 2010). IoT technology involves three layers: the device, which includes hardware, software, and sensors; connectivity, which is how the physical device communicates with the Internet; and the cloud, where data is stored and processed (Wortman & Fluchter, 2015).

The IoT has an assortment of real-world applications to assist users with their everyday lives. These applications can be found in the form of wearables, remote appliance monitoring, smart phones, and smart home assistants (Wortman & Fluchter, 2015). The last device will be the focus of this dissertation. Smart home assistants, in particular, have been developed by many big tech companies (e.g., Google Assistant, Amazon Alexa, iOS Siri, and Microsoft Cortana) and are increasingly being adopted by consumers. It is predicted that by the end of 2023, smart home assistants will become a $8 billion market (Smith, 2018).

Users primarily interact with smart home assistants by giving verbal commands, typically in a task-oriented Q&A format (Cowan et al., 2017). From these verbal interactions, the device can aid users throughout tasks of their everyday lives. That

assistance is made possible from the use of location data and other contextual information (Jennings, 2000; Hauswald et al., 2015). The Google Nest devices, for example, allow users to stream media (e.g., music, news), control their homes (e.g., adjust lighting or thermostat), plan their day (e.g., check traffic, weather, calendar), perform tasks (e.g., create shopping lists, set reminders), communicate with others (e.g., audio and video calls), and get answers to their questions (e.g., report facts and information, perform calculations, translate words across languages). For these features to work, an extensive amount of data about the user must be collected. This poses a challenge to user information privacy, which is influenced by the amount of personal data that the user discloses with the device.

The purpose of this dissertation is to evaluate users' data disclosure behavior with smart homes, identify predictors of those data disclosure decisions, and unify theories that describe user disclosure behavior, specifically the Privacy Paradox, Privacy Calculus, and the Description-Experience Gap.

CHAPTER 2

LITERATURE REVIEW

2.1 Privacy and Security

Whenever someone uses the Internet and the devices connected to it, they leave traces of data about themselves, both intentionally and unintentionally (Hann et al., 2007). Some of this data, such as name, photos, and location, are considered personally identifying information, which can pose risks for privacy infringement. Unlike traditional web browsing, smart homes are equipped with hardware sensors that are always on, collecting data about offline activities that occur in the user's home. That data is then shared to the Internet. Unfortunately, this personal data is often collected and interpreted unknowingly and/or is used for motives unknown to the user, resulting in information privacy infringements (Acquisti & Gross, 2006, 2009; Acquisti & Grossklags, 2004; Christofides et al., 2009; Govani & Pashley, 2005; Gross & Acquisti, 2005; Nov & Wattal, 2009; Tufekci, 2008; Young & Quan-Haase, 2009).

For example, consider the 2018 Facebook-Cambridge Analytica scandal. The company Cambridge Analytica used data mining techniques to develop in-depth user profiles. These profiles were then sold to companies interested in making the most out of targeted advertising (Confessore, 2018). The issue with the firm's actions was their unauthorized data collection from approximately 87 million Facebook users, as reported by Facebook's Chief Technology Officer, Mike Schroepfer (Kang & Frenkel, 2018). That collected data was then sold and abused by politicians world-wide. Major events, including Russian interference in the 2016 US presidential election and spread of

misinformation during the 2016 Brexit campaign, were made possible due to the illegitimate data mining practice of collecting data from users without their permission. This was a violation of personal privacy that led to significant societal outcomes, not just politically, but also an increase in privacy awareness. The scandal brought to light that users, and their data, are treated like a product that is exploited and sold, all without the users' knowledge of it happening. Privacy infringements like these reduce peoples' trust in companies and lead users to believe they have no control over their data (Martin, 2018; Weisbaum, 2018).

Another example of users' loss of data control is the 2017 data breach of the credit bureau Equifax. A total of over 146 million U.S. consumers experienced compromised information regarding their identity. The stolen data included names (146.6 million users), dates of birth (146.6 million users), social security numbers (145.5 million users), phone numbers (20.3 million users), and payment card information (209,000 users; Equifax, 2017). The fact of the matter is that the volume of data generated and collected about users is growing exponentially. This makes it difficult for the user to maintain control of their personal information and for federal legislation to comprehensively regulate how consumer data are treated (Kerry, 2018). Broadly, some users have reported that they feel forced into disclosing their personal data, because they would be unable to use a service if they did not do so.

There are two ways that personal data collected by an IoT device can be abused. The first is from poor security. Adversaries (e.g., unapproved 3rd parties, malicious network observers) can leverage weaknesses in a device's security infrastructure to collect data without the user's knowledge. Security protocols such as data encryption,

two-factor authentication, and the use of passwords help, but they are not always adequate in protecting users' data. Even when encrypted, adversaries can eavesdrop and use inference techniques to identify users' behavior in the confines of their own home (Apthorpe et al., 2017; Chan & Perrig, 2003; Crager et al., 2017). Additionally, digital traces left behind by the user's behavior can be concatenated to develop a single profile on the user and other residents of the household where the smart home assistant is located (Jacobsson et al., 2016). The impacts on the user can range from personalizing services and improving user experience to more severe consequences such as unwanted public disclosure of private behaviors, cyber stalking, burglary by observing when a home is empty, or hijacking control of the home (e.g., recording or eavesdropping on audio and video feeds, changing the temperature or other energy consumption, locking or unlocking smart doors).

The second way that data can be obtained is through poor privacy behaviors exhibited by the user. These behaviors diminish users' control over how and when their personal data is gathered and used. Users often grant a device and/or service access to their data without giving much thought to what that access entails. As technology grows, those requests occur more frequently and for larger quantities of data (Joinson, et al., 2011). The first of the many data requests are in the form of a Privacy Policy and Terms of Service. Unfortunately, these agreements are not always leveraged by the user when making decisions to disclose their data. One reason for this is the length of time required to fully read the agreements. A study on privacy policies for social networking sites by Obar and Oeldorf-Hirsch (2016) reported that 97% of participants agreed to the experimental privacy policy and 93% agreed to the much shorter terms of service.

Although the privacy policy was estimated to take between 29 to 32 minutes to fully read, participants' average reading time was a mere 73 seconds. Similarly, participants spent only 51 seconds on the terms of service, which was estimated to require between 15 to 17 minutes to read. This could indicate that the value in using the device and/or service outweighs the potential risks.

Data privacy is an important piece to user confidentiality, and it is too often violated through means of over disclosure. However, for IoT technologies to work, they require data from the user. The more data the device has, the more it can do for the user. When devices ask for different types of data, users must decide whether or not to disclose that data about themselves. These choices of disclosure are influenced by factors such as privacy concerns, use case value, perceived risks of privacy infringement, and the level of sensitivity that the data possesses. Even when these factors are considered, what users say they will disclose versus what they actually disclose does not always match up.

2.2 Data Disclosure

Much of the privacy literature has shown that users' privacy concerns or intentions to disclose data are inconsistent with what they *actually* disclose (Acquisti & Grossklags, 2005; Cvcek et al., 2006; Taddicken, 2017). This discrepancy is described as a Privacy Paradox (Awad & Krishnan, 2006; Kokolakas, 2017). This holds true for even privacy-conscious individuals (Spiekermann et al., 2001). The Privacy Paradox was first uncovered by Brown (2001) while exploring internet usage while online shopping. It was Brown who observed that although participants had privacy concerns, they still willingly provided their personal information for some sort of gain (e.g., gift

cards, discounts, gifts). The paradox was later more established in a study conducted by Norberg et al. (2007) where the term “Privacy Paradox" was coined. In Norberg’s study, participants indicated their willingness to disclose specific pieces of information in a survey and were later asked to provide this information to a confederate posing as a market researcher. The results supported the notion that people disclose more information about themselves than what they report that they would be willing to. The same conclusions have been found by a number of other studies (Belanger & Crossler, 2011; Smith et al., 2011). Simply put, while people appear to value their information privacy, their actions say otherwise (Nisenbaum, 2009).

Previous research indicates that actual disclosure behavior is directly influenced by self-reported disclosure (Barth & de Jong, 2017; Norberg et al., 2007); however, it has also been shown that other factors influence the mismatch between self-reported disclosure and actual data disclosure. These factors include: privacy literacy (Debatin et al., 2009, Park, 2013; Trepte & Dienlin, 2015), privacy concerns (Acquisti & Grossklags, 2005; Cvcek et al., 2006; Gerber, Gerber, & Volkamer, 2018; Zlatolas et al., 2015), perceived value (Zlatolas et al., 2015), levels of trust in the technology and/or the service provider (Ackerman et al., 1999; Earp & Baumer, 2003; McKnight et al., 2011; Gerber, Gerber, & Volkamer, 2018), and past disclosure (Culnan & Armstrong, 1999). Specifically, data disclosure is higher when users possess low privacy concerns (Krasnova et al., 2009; Wu et al., 2012) or highly value the outcome (Zibuschka et al., 2019). Disclosure is lower when users perceive a higher risk of privacy infringements (Bauer et al., 2016) or receive a request for more sensitive data (Malheiros et al., 2013). Data disclosure is also affected by users’ past disclosure of their data (see Figure 1).

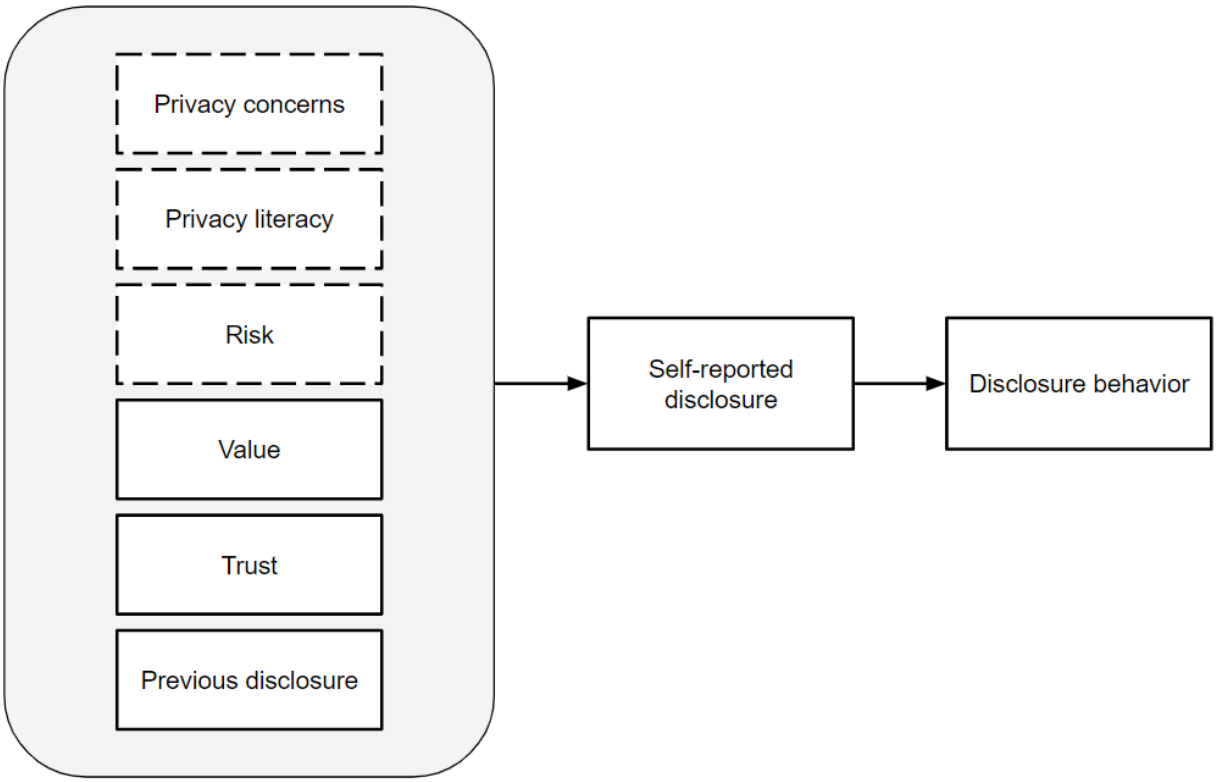


Figure 1

Model of Predictors of Data Disclosure

Note. Solid and dotted lines represent positive and negative predicted relationships, respectively.

2.3 Privacy Concerns

Online, data-driven IoT products and services request that users disclose their personal information for their own benefit (e.g., to save time and money). Greater data disclosure also allows these products and services to compete against other e-commerce entities by personalizing services and implementing targeted advertisements (Chen et al., 2001). To attain these goals, tons of data must be collected about the user and their home, so much so that the devices have the potential to know the user better than their closest friends and family (Jacobsson et al., 2016), which can be concerning to the user.

The popularity of ubiquitous devices, such as smart home assistants, is so pronounced that concerns regarding privacy have become more prominent. People are aware that data, in general, is being collected about them, but they lack clarity about what is collected and why it is used (Cvreck et al., 2006). Specifically, seventy nine percent of Americans are concerned about how much data is collected by companies (Pew Research Center, 2019). Despite the concern, many users admit that they are not regularly attentive to the potentially negative impact of disclosing their information and continue to do so anyway. Even worse, most Americans feel that they cannot even go about their day without having their data collected, resulting in the belief that they have little control over their data (Pew Research Center, 2019) or giving the illusion that they do, in fact, have control over it (Gerber et al., 2018).

One factor that influences privacy concerns is the sensitivity of the data. More sensitive data contains more identifiable information about the user (Malheiros et al., 2013). Perceived sensitivity for different types of personal information widely varies across individuals, influencing peoples' actual disclosure behavior. When data is considered to be highly sensitive, users have higher privacy concerns and lower intentions to disclose that data (Mothersbaugh et al., 2012).

Schomakers et al. (2019) evaluated how people rate the sensitivity of different data types. They found that data that is highly personally identifying, like one's social security number or credit card number, is considered to be highly sensitive information. Phone numbers, locations, pictures of one's face, and voice prints were considered to be moderately sensitive. Low sensitivity data included email addresses and demographic information (e.g., height, hair color, religion, and occupation).

2.4 Risk

When personal data is collected, it can be used to define and trace the owner. This introduces the possibility of data infringements and abuse. Some of these infringements are relatively benign, like using browsing data to generate targeted ads or personalized services for the user (Ur et al., 2012). Alternatively, privacy infringements can be much more malicious in nature, like cyberstalking, identity or financial theft, or violations of medical records. Notably, every two seconds there is a new case of identity theft, and in the United States, 33% of adults have had their identity stolen (Bellemare, 2018).

In the privacy domain, perceived risk is mainly the extent to which users feel that they lack control over their own data. That is, users do not know what or how much data is collected, what that data is used for, and with whom that data is shared. This lack of control can lead to potential privacy infringements. One problem for estimating the potential risk for data disclosure is that the everyday user is not highly literate in aspects of privacy and security. In fact, 59% of Americans report knowing little or nothing about what happens to their data when it is collected by companies and service providers; this proportion is even higher (78%) when the recipient of user data is the government (Pew Research Center, 2019). Even when a company is transparent about their privacy policies, only 20% of Americans report consistently reading them. Further, only 22% of those that do read privacy policies read them completely (Pew Research Center, 2019). Therefore, users' decisions to disclose data are ill-informed because they have not read what their data is used for. When users lack this awareness, they do not know the potential risks of disclosing their data with a device or service provider. Not every user

considers the potential risks of disclosing their data, but when they do, risk perceptions were found to be significantly correlated with intention to disclose information (Naeini et al., 2017). That is, the higher a user perceives their privacy risks to be, the less willing they are to disclose their data. However, risk was not correlated with actual disclosure in the context of marketing research for banks, pharmaceutical companies, health clubs, and reality tv (Norberg et al., 2007).

The choices that people make, including those relating to data disclosure, have associated risks. The way that people interpret these risks is affected by context, as described by the Description-Experience Gap (Camilleri & Newell, 2013; Hadar & Fox, 2009). When decisions are made from description, users have on-hand information about the consequences of their actions and the probability of these consequences occurring (Newell & Rakow, 2007). In the case of data disclosure, a user might learn that roughly three in ten Americans have experienced some sort of privacy infringement in 2019 (Pew Research Center, 2019). This information could directly inform their self-reported intentions to disclose. However, in real-life disclosure situations, this descriptive information is not often available. In these circumstances, individuals must rely on their experiences (Hertwig & Erev, 2009), which involve uncertainty (Schomakers et al., 2019). The measurement of decisions of experience as described in the Description-Experience Gap will be referred to as “perceived risk” in this study. Numerical risk information will not be presented in the research study, so perceived risk will be the focus. Therefore, respondents’ disclosure decisions are based on their perceptions of risk. It is possible that users’ data disclosure differs in the context of

quantified risks; this lies beyond the scope of this dissertation but provides an important direction for future research.

2.5 Value

In the economic literature, users' personal data is thought to have monetary value. Some research has looked at how much users would expect to be paid for their data (Acquisti et al., 2013; Steinfeld, 2015), while other researchers looked at how much how much users would be willing to pay to protect their privacy (Mihale-Wilson et al., 2017). When individuals earn some form of monetary reward (e.g., discounts, gifts), they are more willing to trade their personal information. This was found to be true even for those that reported having high levels of privacy concerns (Spiekermann et al., 2001).

There is, however, another perspective of value when it comes to privacy -- the value of the service itself and how it influences user disclosure behavior (i.e., utility). Users perform a "Privacy Calculus" by evaluating the perceived risks and benefits of disclosing their personal information. This evaluation leads to a decision about whether or not to disclose that data (Dinev & Hart, 2006; Laufer & Wolfe, 1977). The benefits of disclosing information include convenience and personalization of services (Hann et al., 2002), social rewards (Krasnova et al., 2012), and monetary incentives like discounts (Carrascal et al., 2013; Sayre & Horne, 2000; Steinfeld, 2015). Although 81% of Americans report that the potential privacy risks outweigh the benefits of disclosing their data (Pew Research Center, 2019), their behavior says otherwise. The disparity between willingness to disclose data and actual disclosure behavior, as depicted by the

Privacy Paradox, can be further explained by evaluating users' Privacy Calculus (see Figure 4).

Perceived value can also be defined in terms of perceived usefulness. In the Technology Acceptance Model (TAM), proposed by Davis (1985), users' acceptance, perceived ease of use, and perceived usefulness of a device informs their intentions to use it (see Figure 2). It is similar to previous literature such that intentions directly inform a person's actual behavior. For example, Davis & Venkatesh (1996) showed that perceived usefulness directly leads to one's intentions to use a device, which then leads to one's actual usage. This flow of behavior has similarities to models of data disclosure behavior where disclosure intentions directly inform actual behavior (Norberg et al., 2007; Gerber et al., 2018). Although TAM has to do with usage, the architecture of the model mirrors the way that empirical research illustrates data disclosure behavior.

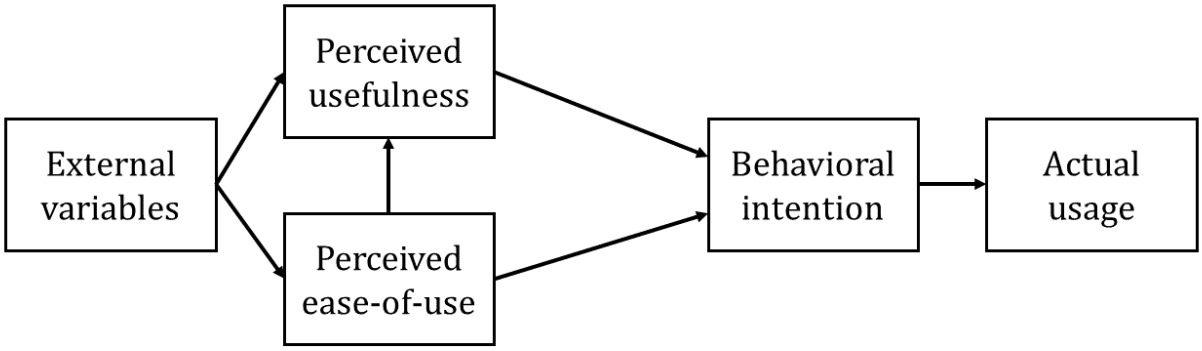


Figure 2
Technology Acceptance Model (Davis & Venkatesh, 1996)

Rewards tend to be weighted so heavily that sometimes users do not hold any privacy concerns for their data at all (Park et al., 2012). In other words, users place more emphasis on reward than on their privacy concerns, which then impacts their

disclosure decisions. One explanation for the biased weighting of rewards is due to delay discounting, where short-term rewards are valued more than one's future privacy (Acquisti, 2004; Acquisti & Grossklags, 2005). For instance, if an individual requires directions to an unknown location, their location data will be required for the IoT to compute the best route to take. The short-term need to get from Point A to Point B can be valued more heavily than long-term concerns about sharing one's location data. Not only does experienced value influence data disclosure behavior, but expectations of future benefit also influence what and how much data users share (Hann et al., 2007; Vroom, 1964).

2.6 Trust

One of the most significant components to consumer tech adoption and data disclosure is trust (Castelfranchi & Falcone, 2005). When users have trust in an entity, they expect that it can be relied on to do exactly what it was intended to do and to act in the user's best interests. Trust is central to decisions that have any uncertainty or potential risks involved (Hertwig et al., 2004). The more trust a user has in a given technology, the more likely they are to adopt it and to disclose their data to it (Lankton et al., 2014). Trust in the context of data disclosure can come in two forms: one's propensity to trust (Tait & Jeske, 2015) and one's trust in the technology and/or vendor itself (Chung et al., 2017). This dissertation does not distinguish between these forms of trust because such a distinction would require evaluating individual differences to determine users' propensity to trust in entities at multiple levels of the IoT. For this reason, the measurements of actual disclosure and self-reported disclosure in this dissertation are not contextualized to be company- or personality-specific, allowing

individual differences to be captured in the error variance (i.e., standard errors of the estimates). The purpose of this dissertation is to create a holistic model of user data disclosure decisions; subsequent studies can and should be used to evaluate individual differences as moderators of this relationship.

The Privacy Calculus model suggests that users evaluate both the costs and benefits of their data disclosure. Trust is a piece of this decision-making process (Hoffman et al., 1999). When users do disclose their data, they trust that their data is safe. Metzger (2004) conducted a study to understand what cultivates trust in data disclosure. The study found that: (1) trust is positively correlated with data disclosure to a website and (2) privacy concerns, perceived security, and opinions of the company affect users' sense of trust for that website. This indicates that the factors (e.g., privacy concerns, perceived risk) that influence data disclosure behavior are also interrelated. Noticeably, the perceived risks of data disclosure are lessened when a user trusts the technology (Rubin, 1975).

2.7 Privacy Models

Three other noteworthy privacy models describe the ways that users and developers think about and understand privacy as a construct (Mai, 2020). The Panopticon Model proposes that people think of privacy as a form of paranoia, or the feeling that someone, or something, is always watching (Campbell & Carlson, 2002). The mindset described by the Panopticon Model is uniquely descriptive of bad actors that seek to steal user data without explicit permission granted by the data's owner. Negative outcomes like these are captured within the context of Privacy Calculus in the form of user's sense of risk, their privacy concerns that their data can and/or will be

abused, and their trust by acknowledging there are always some weakness in technology that can result in data leakage.

The Capture Model speaks to the aftermath of data disclosure, specifically how this data is coded, or interpreted, by technology (Agre, 1994). There is always a possibility that once data is obtained by a recipient (e.g., product company) or device, it is misused beyond what the user intended to disclose their information for. For example, if a person were to share their phone number with a smart home, the product company may sell that information to another recipient so that they can contact the person unwarranted. This model is a metaphor to describe the risks of user data disclosure, but not how it impacts their disclosure decisions.

The Datafication Model suggests that people and developers think of privacy in the context of how data is collected and used (Mai, 2016). Similar to the Panopticon and Capture models, this describes a mindset people use when thinking about privacy abstractly, rather than a framework for understanding specific factors that inform individual users' data disclosure decisions.

The purpose of this study was to understand the predecessors of data disclosure decisions, rather than how users and developers think about and understand these disclosure behaviors after they occur. Therefore, the Privacy Calculus model was chosen over the other models for further study.

2.8 Weaknesses of the Existing Literature

The most glaring gap in the privacy and data disclosure literature is the lack of direct observation of data disclosure. That is, existing studies may not accurately reflect users' actual data sharing behaviors because they are based on one's self-reported

intentions to disclose (Camilleri & Newell, 2013; Hau et al., 2008; Zeng et al., 2017). A limited number of studies suggest there is a mismatch between users' self-reported disclosure and their actual behavior (e.g., Beresford et al., 2012; Egelman et al., 2012; Hann et al., 2007, Huberman et al., 2005; Norberg et al., 2007; Spiekermann et al., 2001). However, these studies did not evaluate factors other than privacy concerns that may predict disclosure behavior.

The literature regarding the Privacy Paradox is also inconsistent. Some researchers have found evidence in support of the phenomenon (Barth & de Jong, 2017; Gerber et al., 2018), while others have not (Hughes-Roberts, 2013; Taddicken, 2014; Tufecki, 2008). This is possibly due to the fact that the studies have evaluated the moderating factors (e.g., privacy concerns) of disclosure behavior separately, but not together. This includes privacy concerns, perceived risk, perceived data sensitivity, and value of different uses for devices. Furthermore, little to no research has evaluated these factors in relation to actual disclosure behavior.

In the current Privacy Paradox literature, studies have observed that despite people's value in privacy, they do not always behave in a way that supports their position on the importance of protecting their personal data. However, this does not provide a full or accurate picture as to why participants disclose data in the way that they do. That is, disclosure behavior is influenced by more than a person's willingness to disclose their data. Privacy Calculus suggests that users weigh the perceived costs against the perceived benefits of data disclosure to decide whether or not to share their data. It is possible that differences in the weighting of these factors are responsible for the misalignment between self-reports and behavioral measurements of data

disclosure. Therefore, it is possible that Privacy Calculus has an impact on the Privacy Paradox and can provide a deeper explanation as to why there is a discrepancy between self-reported disclosure and actual data disclosure behavior (see Figure 3). Unfortunately, there is little unification between the two theories in the data disclosure literature, which emphasizes the need to evaluate disclosure behavior in conjunction with factors that influence decisions on disclosing one's data.

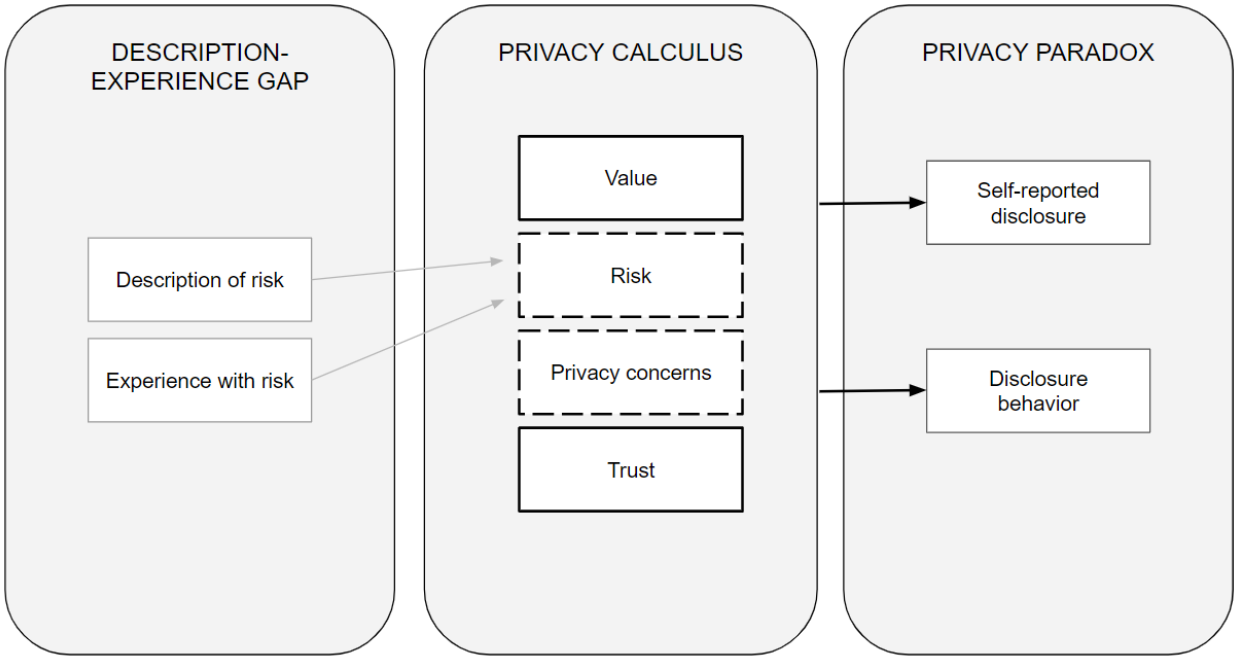


Figure 3

A Proposed Model to Unify Theories Relating to Data Disclosure

Note. Boxes with solid and dotted borders represent positive and negative relationships to data disclosure decisions, respectively. Lighter gray arrows from the Description-Experience Gap were not explicitly evaluated but are acknowledged due to evidence on how risk is evaluated.

Another weakness that exists in the literature relates to generalizability. Privacy behavior is contextual in nature (Morando et al., 2014), but most research on privacy behaviors is contextualized in e-commerce and social networking (Barnes, 2006;

Beresford et al., 2012; Brown, 2001; Huges-Roberts, 2013; Taddicken, 2014; Tufekci, 2008). Therefore, those findings may not generalize to other environments. This dissertation addresses a specific gap in the literature by testing the extent to which the Privacy Paradox and Privacy Calculus can account for data disclosure behavior as it relates to assistant technology.

Finally, the discrepancy between user's self-reported and actual disclosure indicates a weakness in the external validity of self-report measures. There has been an overwhelming use of survey and interview data to evaluate disclosure behavior, however people do not always behave the same way they say they would. Surveys and interviews are appropriate for examining general privacy attitudes and concerns, but they do not provide a full picture of actual disclosure behavior (Kokolakis et al., 2017). Although there are a handful of studies that did administer experimental tasks (Beresford et al, 2012; Egelman et al., 2012; Hann et al., 2007, Huberman et al., 2005; Norberg et al., 2007; Spiekermann et al., 2001), these studies did not evaluate the relationship between intentions to disclose, actual disclosure, and antecedents (e.g., privacy concerns, perceived risk, data sensitivity, perceived use case value, trust) that may influence disclosure behavior. That is, the Privacy Paradox and Privacy Calculus models have not been completely unified in the context of smart home devices.

2.9 Purpose

The purpose of this research was to unify the Privacy Paradox and Privacy Calculus Model by answering two questions: (1) how does self-reported disclosure relate to actual disclosure behavior, and (2) what factors predict disclosure decisions of both kinds? The hypotheses tested are:

H1: Users will behaviorally disclose more data than what they report intending to disclose.

H2: Participants will express greater intentions of disclosing their data and behaviorally disclose more data when the value of the use case is rated higher.

H3: Participants will express lower intentions of disclosing their data and behaviorally disclose less data when their perceived risk for disclosing a specific data type is higher.

H4: Participants will express lower intentions of disclosing their data and behaviorally disclose less data when their privacy concerns are higher.

H5: Participants will express lower intentions of disclosing their data and behaviorally disclose less data when their level of trust is higher.

These hypotheses were first addressed within the context of a pilot study that explored the relationship of value, risk, privacy concerns, trust, and data sensitivity on people's self-reported disclosure decisions. They were then confirmed within a three-part follow-up study that measured both self-reported disclosure and actual behavioral disclosure.

CHAPTER 3

PILOT

3.1 Purpose

Participants completed an online survey asking questions of their self-reported data disclosure. The data from this study were used to calculate the appropriate sample size for the analyses proposed in Chapter 4.

3.2 Method

Forty-three participants (31 females, M_{age} = 20.7, SD_{age} = 8.11, 58.1% smart home owners) were recruited through Wichita State University's SONA systems. These participants were granted SONA credits required for Psychology courses. To participate, participants had to be at least 18 years of age, be a fluent English speaker, and own a smartphone.

Participants answered demographic questions on age, gender, and smart home ownership. The remainder of the survey consisted of a series of vignettes that asked whether the participant would disclose a particular type of data in each situation (i.e., use case). Five different use cases (e.g., check the weather, make an audio call,) and five different data types (e.g., location, microphone access) were referenced (see Appendix C). These use cases and data types were used to inform the Device Set Up and Phone Tasks in the subsequent, primary experiment. All combinations of use case and data type were presented for a total of 25 vignettes. The vignettes were also device neutral, meaning no specific devices were mentioned. An example vignette is as follows:

*A device requests access to your **camera** to **view a live video feed** of the room in your house. Will you provide this information, yes or no?*

After reading each vignette, participants indicated whether they would disclose information by selecting “Yes” or “No”. Participants then answered four follow-up questions:

How valuable is being able to view a live video feed of a room in your house?

How risky is it to you to grant access to your camera?

How sensitive, or private, is your camera data (i.e., photos, videos) to you?

How concerned are you about disclosing your camera data (i.e., photos, videos)?

Participants responded to each question on a scale from 1 to 7. Scale anchors depended on question type: use case value (*not at all valuable* to *extremely valuable*); data type risk disclosure (not at all risky - extremely risky); data type sensitivity (not at all sensitive - extremely sensitive); and concern with data disclosure (not at all concerned - extremely concerned).

Vignettes were presented one at a time, in a randomized order. After each vignette, participants responded to the four follow-up questions on value, risk, data sensitivity, and privacy concerns. Therefore, the value, risk, sensitivity, and concern ratings were requested a total of 25 times. Participants did not provide specific personal information, only their intentions to disclose that data in the various contexts. The use cases included in the survey included check the weather, find phone, make an online

purchase, place an audio call, and view a live video feed. The data types included were location, phone number, credit card information, microphone, and camera.

3.3 Results

A logistic regression was conducted to assess the impact of perceived risk, perceived data sensitivity, and privacy concerns on participants' data self-reported disclosure. The regression also included the main effects and interaction between use case and data type to control for context effects. The model was statistically significant, χ^2 (27, N = 1074) = 591.696, $p < .001$. The hypothesized predictors were successful in differentiating those that did intend to disclose their data from those that did not. Specifically, the predictors explained 54.4% of the variability in participants' self-reported disclosure (Nagelkerke R^2), with 79.5% of cases correctly classified. The odds ratios for value, risk, and concern were 2.31, 0.52, and 0.99, respectively.

The model indicated that value and risk contributed significantly to participants' disclosure decisions (Wald χ^2 (1, N = 1074) = 66.12, $p < .001$; Wald χ^2 (1, N = 1074) = 17.44, $p < .001$, respectively). On average, participants rated their perceived value of the use cases at 4.75; at this value they were 17.4% likely to disclose their data. In contrast, when participants highly valued the use case (i.e., value = 7), they became 36.5% likely to disclose their data. In other words, there is a positive relationship between self-reported disclosure and use case value, where the higher valued use cases resulted in a higher likelihood of data disclosure.

On average, participants rated their perceived risk of data disclosure at 4.84; at this value they were 17.2% likely to disclose their data. In contrast, when participants perceived disclosure risk to be low (i.e., risk = 1), the likelihood of data disclosure was

47.0%. In other words, there was a negative relationship between self-reported disclosure and perceived risk, where the lower perceived risk resulted in a higher likelihood of data disclosure. Privacy concern did not significantly contribute to participants' disclosure decisions (Wald χ^2 (1, N = 1074) = .006, p = .940).

Use case (Wald χ^2 (4, N = 1074) = 37.97, p < .001), data type (Wald χ^2 (4, N = 1074) = 36.48, p < .001), and their interaction (Wald χ^2 (16, N = 1074) = 156.126, p < .001) also significantly contributed to participants' disclosure decisions. These variables were included in the model to control for data type and use case characteristics that may affect disclosure decisions; however, these effects were not hypothesized and will not be discussed further.

3.4 Power Analysis

The logistic regression output was used to conduct a power analysis in G*power 3.1 (Faul et al., 2007, 2009; Mayer et al., 2007) to determine a sufficient sample size for each of the predictor variables: perceived value, perceived risk, data sensitivity, and privacy concerns. Input parameters to reach the recommended sample size included: two-tailed, normal distribution, specific predictors' odds ratios, a .05 alpha value, and a power of .80 (as recommended by Cohen, 1992). For the perceived value, perceived risk, and privacy concern predictor variables, the recommended sample sizes were 88, 135, and 276,879, respectively. The desired sample size was set at 90 participants to achieve sufficient power while compromising to keep the study feasible.

CHAPTER 4

METHOD

4.1 Participants

Ninety-one individuals (64 female, 27 male) participated in this study. The average age was 22.9 years *(SD* = 8.11*)*. There were 56 participants that were smart home owners representing 61.5% of the sample size; 35 were non-owners. Seventy-two participants (79.1%) were iOS users and 19 (20.9%) were Android users.

Participants were recruited through Wichita State University's SONA system and externally shared advertisements and were compensated with 6 SONA credits or a $15 gift card. People were required to be at least 18 years of age, own a smartphone, and speak fluent English to be eligible to participate.

4.2 Procedure

This study consisted of three experimental tasks that were presented in a counter-balanced order — a vignette survey, a smart home interaction task, and a phone disclosure survey — and concluded with a short debrief survey. Sessions took place remotely and were held over Zoom to maximize the health and safety of participants.

4.2.1 Vignette survey

The vignette survey measured participants' willingness, or intention, to disclose their data. A total of 56 scenarios (vignettes) were presented in which participants chose "Yes" or "No" to indicate whether they would disclose their information to a smart home device in specific situations. Each Vignette included one use case and one data type pulled from a total of seven different use cases and eight different data types (Table 1).

Table 1
Vignette Survey Use Cases and Data Types

Use cases	Data types
Check the local weather	Location / zip code
Find a misplaced phone	Phone number
Make an online purchase	Credit card information
Make an audio/phone call	Microphone access
Check a live video feed	Camera access
Set a calendar event	Calendar access
Stream media	Photo access
	3rd party app account access

To minimize any potential order effects, presentation order of the Vignettes was randomized across participants. After each Vignette (i.e., use case and data type pairing), participants responded to four 7-point Likert scale questions pertaining to use case value, privacy risk, privacy concerns, and level of trust.

Example Vignette:

A smart home requests access to your camera to view a live video feed of the room in your house. Will you provide this information, yes or no?

- How valuable is being able to view a live video feed of a room in your house?
    - 1 (*not at all valuable*)
    - 7 *(extremely valuable*)
- How risky is it to you to grant access to your camera?
    - 1 (*not at all risky*)
    - 7 (*extremely risky*)
- How concerned are you about disclosing your camera data (i.e., photos, videos)?
    - 1 (*not at all concerned*)

    - 7 (*extremely concerned*)
- How trustworthy do you feel smart home assistants are in keeping your camera data private?
    - 1 (*not at all trustworthy*)
    - 7 (*extremely trustworthy*)

At the very end of the survey, participants responded to a series of open-ended questions expanding on why they rated use case value, perceived risk, privacy concerns, and trust in the ways that they did. The full survey can be located in Appendix A.

4.2.2 Smart home interaction task

The smart home interaction task measured participants' actual data disclosure behavior. Participants were presented with an interactive prototype of a smart home device created in Axure RP 10 (Schwartz & Srail, 2014). Participants used this prototype to complete five sub-tasks in a counterbalanced order. Each sub-task represented a unique feature that a smart home can be used for (i.e., use case) and had an associated data type that participants were asked to provide. The tasks assigned to participants (see below) were designed to maximize external validity to smart home usage. For example, a smart home is likely to request the user's zip code to check the weather, but it will not request a person's credit card number.

The prompts and associated data requests were as follows:

- Check the weather, zip code
- Find your phone, phone number
- Make an online purchase, credit card number

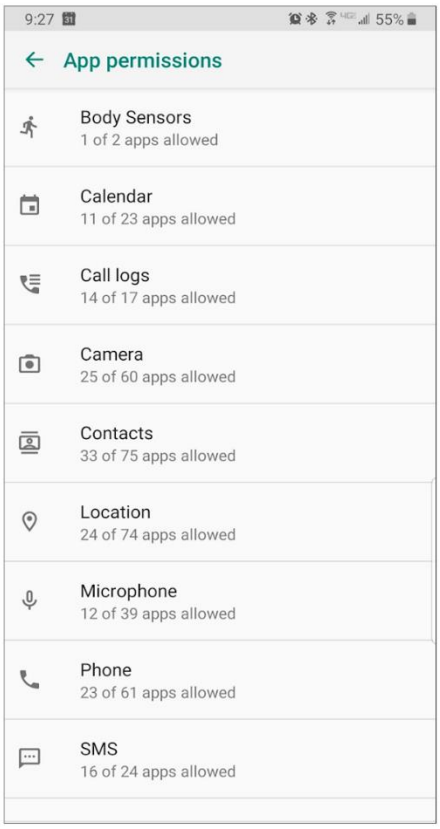
- Place an audio call, mic access
- View a live video feed of your home, camera access

Binary “Did” or “Did not” disclose information was recorded by the researcher as participants completed each sub-task. Each prompted data type was optional for participants to provide.

4.2.3 Phone survey

A phone survey was also conducted as a precautionary measure in case of prototype failure or data collection difficulties due to COVID-19. Participants provided the total number of mobile apps that requested access to their phone and the number of apps that they granted access to their personal calendar, camera, location, microphone, and phone data. This data was converted to a proportion (total granted access/total requested) to control for differences in the number of apps participants had on their phones. The proportion of apps granted access to each data type requested access was recorded to prevent skewing from calculating raw averages.

Android



iOS

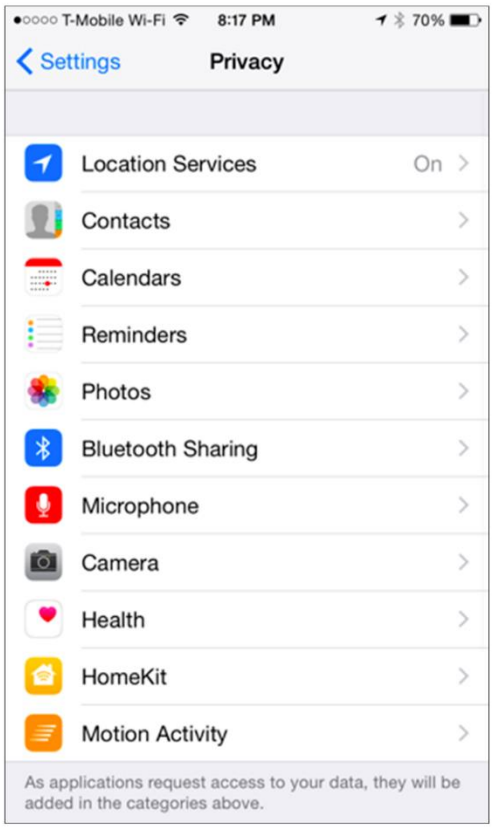


Figure 4

App Permissions Folder on iOS and Android Smartphones

CHAPTER 5

RESULTS

Primary analyses consisted of three individual multi-level logistic regressions to test Hypotheses 1-5. A thematic analysis was conducted on the qualitative data collected from the Vignette Survey. An exploratory analysis on ownership was also performed with a multi-level logistic regression for both self-reported and actual disclosure behavior.

5.1 Privacy Paradox

A multi-level logistic regression (Gelman & Hill, 2006) was used to determine the degree to which participants' self-reported disclosure aligned with their actual disclosure in the Smart Home Interaction Task using the lme4 package (Bates et al., 2020) in R (R Core Team, 2020). The fixed-effect structure included self-reported disclosure, which was effect coded prior to analysis. The random effect structure allowed the intercept to vary across participants and use case to account for individual differences in people's general disclosure behavior and in the appropriateness of disclosure across use cases.

The results of the multi-level logistic regression showed a statistically significant relationship between self-report and behavior; however, the effect size was quite small (see Table 3). Those who reported they would not disclose their data were 94% likely to behaviorally disclose when faced with the real-world context, while those that self-reported "Yes" were 99% likely to behaviorally disclose under the same conditions. Therefore, self-reported disclosure is not a strong predictor of disclosure behavior, and the Privacy Paradox for smart home devices is supported (Hypothesis 1).

Table 2
Summary of Analyses

Analysis	Hypotheses	Statistical approach
Privacy Paradox	H1	Multi-level logistic regression *Self-report ~ behavior*
Privacy Calculus: Self-reported disclosure	H2 H3 H4 H5	Multi-level logistic regression *Self-reported ~ value + risk + concern + trust*
Privacy Calculus: Disclosure behavior	H2 H3 H4 H5	Multi-level logistic regression *Behavior ~ value + risk + concern + trust*
Reasoning for each predictor rating	H2 H4 H5	Thematic analysis Percent of coder agreement Krippendorff's alpha
Exploratory analysis of ownership	N/A	Multi-level logistic regression *Self-reported ~ value + risk + concern + trust + ownership + (value*ownership) + (risk*ownership) + (concern*ownership) + (trust*ownership)* *Behavior ~ value + risk + concern + trust + ownership + (value*ownership) + (risk*ownership) + (concern*ownership) + (trust*ownership)*

Note. Predictors are signified with a "~" and main effects are symbolized with a "+" sign.

Table 3

Relationship Between Self-reported and Behavioral Disclosure Decisions

	B	*SE*	*z*	*p*
Intercept	3.63	.88	4.10	< .001
Self-reported disclosure	-.86	.29	-3.00	< .01

Note. Self-reported disclosure was effect coded, with did disclose serving as the {-1} baseline.

5.2 Predictors of Self-reported Disclosure

A multi-level logistic regression was conducted to determine which variables in the Privacy Calculus model predicted participants' self-reported disclosure. The fixed effect structure included the main effects of value, risk, concern, and trust. The intercept was included in the random effect structure to control for individual differences across participants, use case, and data type. All variables were means-centered prior to analysis.

The model indicated that there were significant main effects of value, privacy concerns, and trust supporting Hypotheses 2, 4, and 5. Risk was not a statistically significant predictor, refuting Hypothesis 3 (see Table 4). A VIF test confirmed that the predictors were not multicollinear (value = 1.01; risk = 2.65; concern = 2.70; trust = 1.22), ruling out the possibility that these variables were overlapping constructs.

Table 4

Parameter Estimates from a Multi-level Logistic Regression of Self-Reported Disclosure

	B	*SE*	***z***	*p*
Intercept	-0.43	0.39	-1.10	.27
Value	0.23	0.03	7.87	< .001
Risk	-0.04	0.05	-0.60	.55
Privacy concerns	-0.57	0.05	-9.85	< .001
Trust	0.25	0.05	4.80	< .001

Note. Value (*M* = 4.29), Risk (*M* = 4.10), Privacy concerns (*M* = 3.57), and Trust (*M* = 4.00) were means-centered.

Higher value ratings resulted in a greater likelihood of self-reported disclosure. For value ratings of 7 (*extremely valuable*) participants were 76.59% likely to self-

reported disclose while value ratings of 1 (*not at all valuable*) had a 45.57% likelihood of self-reported disclosure (see Figure 5).

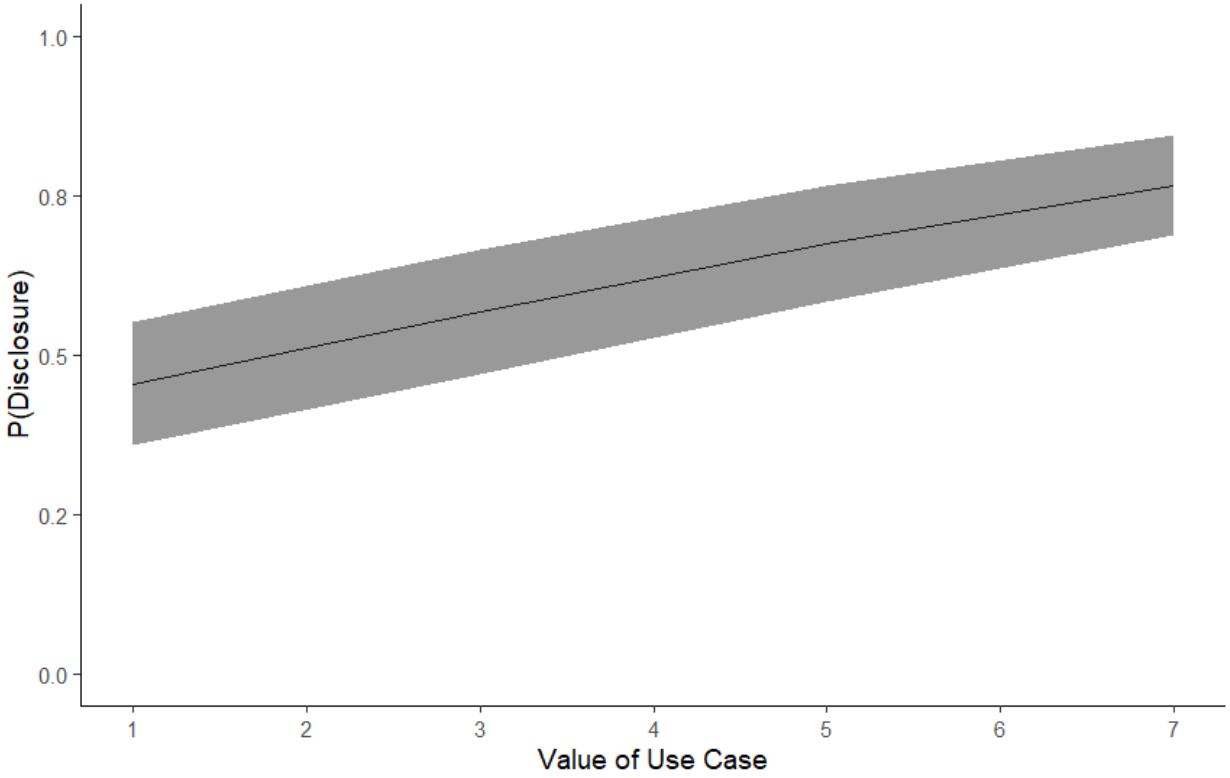


Figure 5

Main Effect of Value on Self-Reported Disclosure

Note. Error ribbon represents ±1 SE.

Higher ratings of trust resulted in a greater likelihood of self-reported disclosure. For trust ratings of 7 (*extremely trustworthy*), there was a 79.26% likelihood of self-reported disclosure and a 46.11% likelihood of disclosure when trust was rated as a 1 (*not at all trustworthy*; see Figure 6).

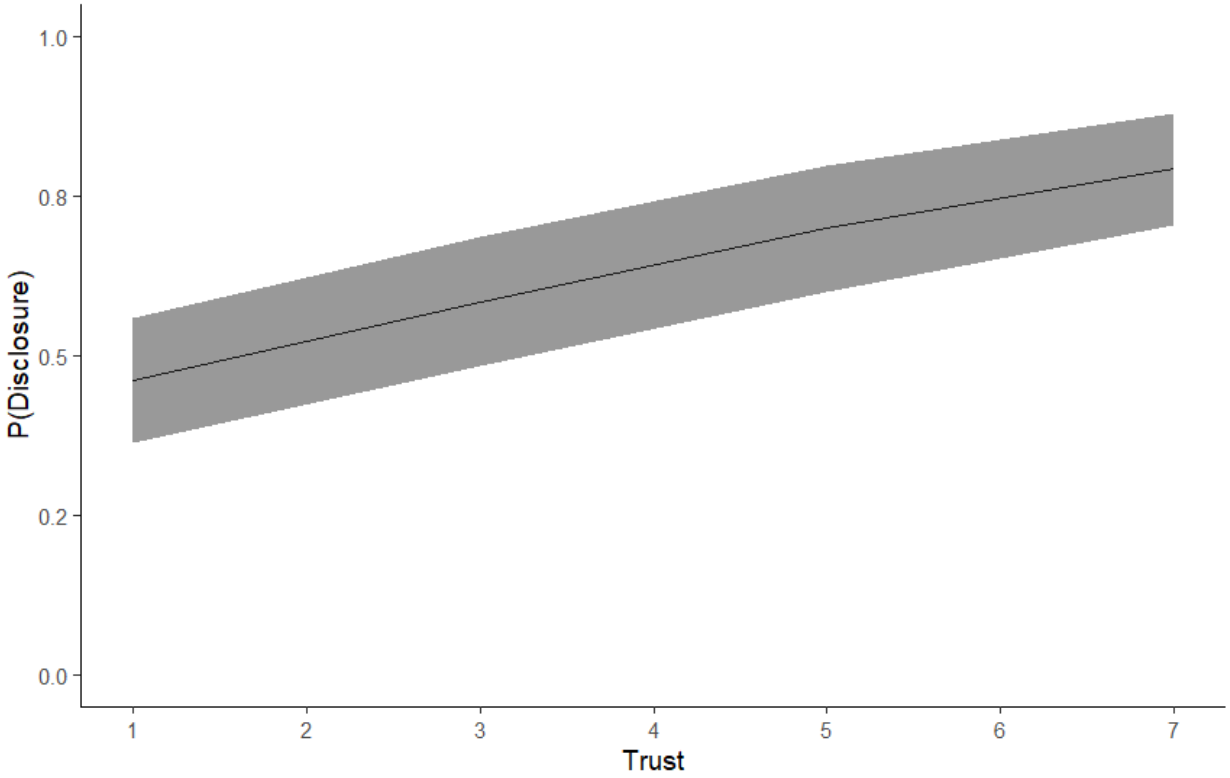


Figure 6

Main Effect of Trust on Self-Reported Disclosure

Note. Error ribbon represents ±1 SE.

Greater privacy concerns value resulted in lower self-reported disclosure. For concern ratings of 1 (*not at all concerned*), there was a 26.96% likelihood of self-reported disclosure and a 1.16% likelihood of disclosure when concern was rated as a 7 (*extremely concerned*; see Figure 7).

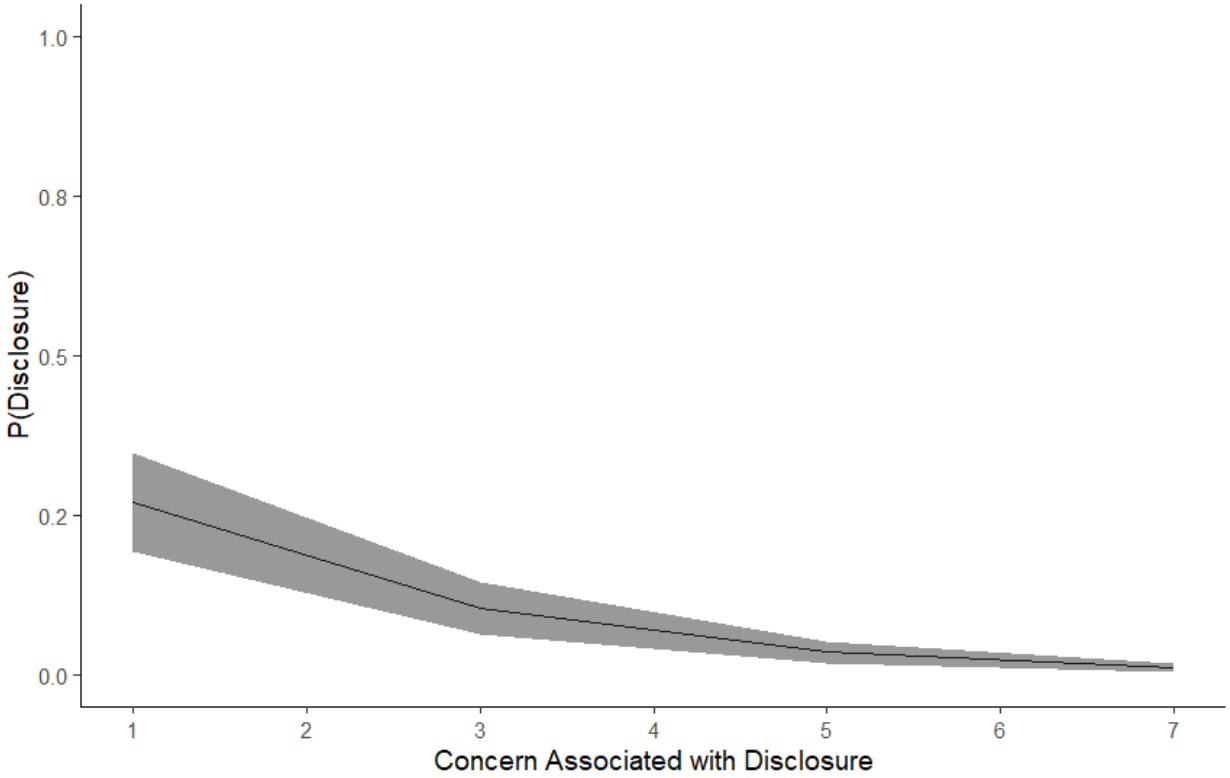


Figure 7

Main Effect of Concern on Self-Reported Disclosure

Note. Error ribbon represents ±1 SE.

5.3 Predictors of disclosure behavior

A multi-level logistic regression was conducted to determine which variables in the Privacy Calculus model predicted participants' disclosure behavior. The fixed effect structure included the main effects of value, risk, concern, and trust. The intercept was included in the random effect structure to control for individual differences across participants and use case. Data type was not included in the intercept random effect structure because data type was perfectly correlated with use case (i.e., there was only one data type for each use case).

The results of the multi-level logistic regression showed a main effect of value supporting Hypothesis 2. Risk, concern, and trust were all insignificant, refuting Hypotheses 3, 4, and 5 (see Table 5). A VIF test confirm that the predictors were not multicollinear (value = 1.82; risk = 3.04; concern = 2.80; trust = 2.14), confirming that these variables were treated as separate constructs by participants.

Table 5

Parameter Estimates from a Multi-level Logistic Regression Predicting Disclosure Behavior

	B	*SE*	*z*	*p*
Intercept	6.60	2.28	2.56	.01
Value	0.70	0.32	2.16	.03
Risk	−0.83	0.54	−1.54	.12
Privacy concerns	−0.33	0.49	−0.68	.50
Trust	−0.18	0.37	−1.31	.19

Note. Value (*M* = 4.40), Risk (*M* = 4.17), Privacy concerns (*M* = 3.64), and Trust (*M* = 3.90) were means-centered.

Although the model indicated a significant effect of value, most participants (N = 91) disclosed their data nearly every time it was requested during the smart home interaction task (see Table 6). Consequently, for each value rating level (1–7) there was a 99.9% likelihood of participants behaviorally disclosing data, which is why the slope of this main effect is flat (see Figure 8).

Table 6

Percent of Cases in Which Data was Disclosed

Data type	*M*	*SD*
Location	99%	10.5%
Phone number	98%	14.7%
Credit card number	68%	46.9%
Microphone	97%	18.0%
Camera	90%	30.0%

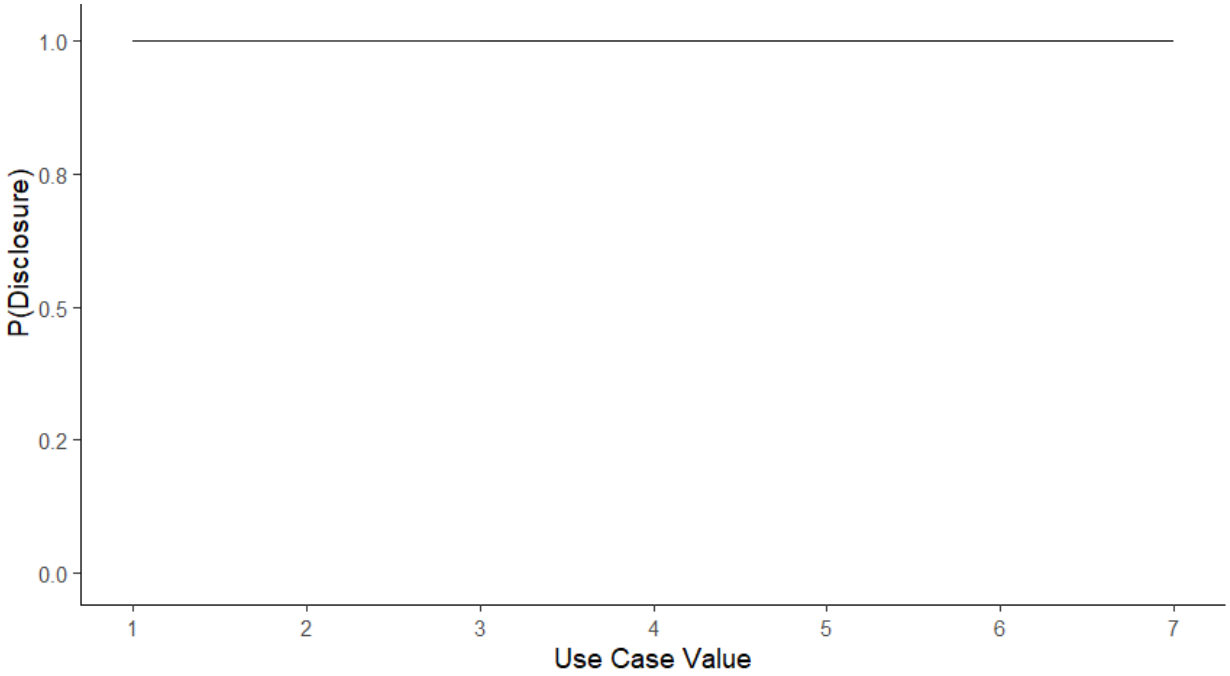


Figure 8

Main Effect of Value on Disclosure Behavior

Note. Error ribbon represents ±1 SE.

In summary, The Privacy Paradox was supported. Value, trust, and concern were significant predictors of self-reported disclosure. Value was the only significant predictor of disclosure behavior (for a summary, see Figure 9).

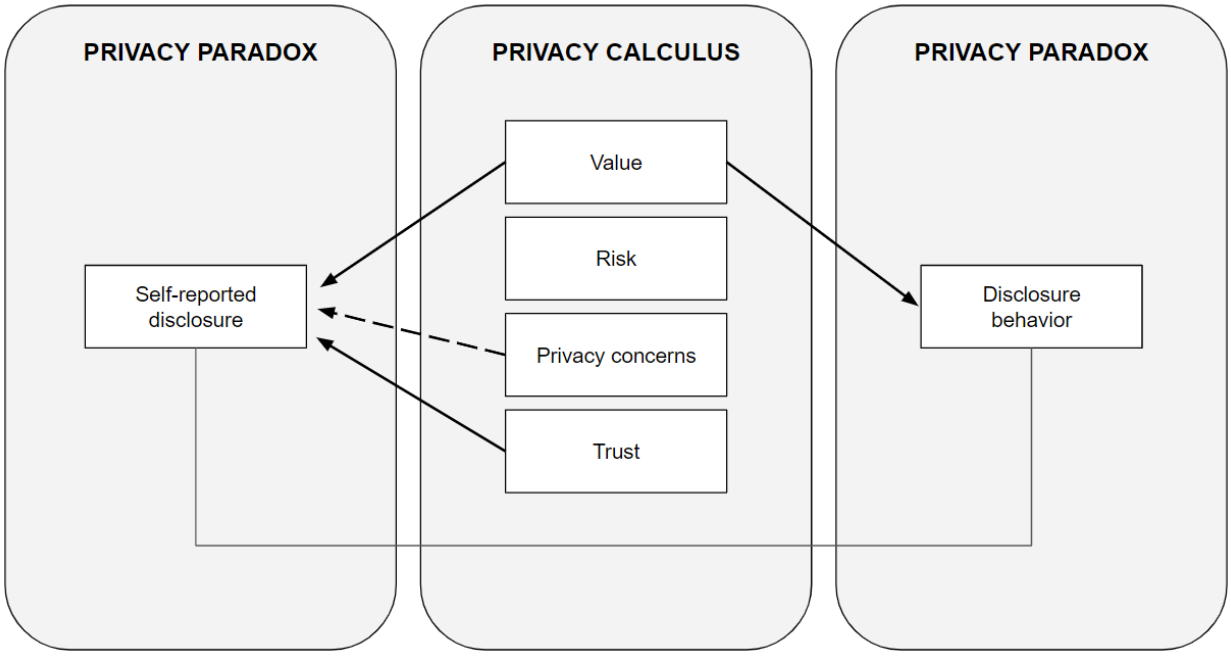


Figure 9

Summary of Results

Note. Arrows represent predictors of self-reported disclosure and behavior. Dotted arrows represent a negative relationship and solid arrows represent a positive relationship.

5.4 Exploratory Analyses of Smart Home Device Ownership

Two multi-level logistic regressions were used to determine whether participants' ownership status of smart home devices impacted their self-reported disclosure and/or disclosure behaviors. The models for both self-reported disclosure and disclosure behavior included main effects of value, risk, privacy concerns, trust, and ownership. The models also included the Value × Ownership, Risk × Ownership, Privacy Concerns × Ownership, and Trust × Ownership interactions. As in the primary analyses, the intercept for each of the two models was also set to vary across participants and use

case. The model for self-reported disclosure also allowed the intercept to vary across data type.

Ownership did not impact participants' self-reported disclosure (see Table 7). A Tukey's HSD post-hoc analysis, conducted with the emmeans package (Lenth, 2019), confirmed that there was no significant main effect of ownership (p = .68) nor any significant interactions of ownership with value (p = .25), risk (p = .75), privacy concerns (p = .93), or trust (p = .27). Additionally, the strength of the relationship between the existing predictors and self-reported disclosure did not significantly change (see Table 7).

Table 7

Impact of Ownership on Self-reported Disclosure

	B	*SE*	*z*	*p*
Intercept	−0.44	0.39	−1.15	.25
Value	0.24	0.03	7.71	<.001
Risk	−0.03	0.06	−0.51	.61
Concern	−0.58	0.06	−9.73	<.001
Trust	0.26	0.05	4.90	<.001
Ownership	−0.07	0.06	−0.42	.67
Value × Ownership	0.03	0.03	1.15	.25
Risk × Ownership	0.02	0.06	0.61	.75
Privacy concerns × Ownership	0.01	0.06	0.09	.93
Trust × Ownership	0.06	0.05	1.10	.27

Note. Value (M = 4.29), Risk (M = 4.10), Privacy concerns (M = 3.57), and Trust (M = 4.00) were means-centered. Ownership was effect coded with owner serving as the {−1} baseline.

Ownership also did not impact participants' disclosure behavior (see Table 8). There was no significant main effect of ownership (p = .62) nor any significant interactions of ownership with value (p = .32), risk (p = .66), privacy concerns (p = .43),

or trust (p = .76). Additionally, the strength of the relationship between the existing predictors and self-reported disclosure did not significantly change (see Table 8).

Table 8

Impact of Ownership on Disclosure Behavior

	B	*SE*	*z*	*p*
Intercept	6.06	1.98	3.06	<.01
Value	0.69	0.28	2.43	.02
Risk	-0.84	0.52	-1.62	.11
Concern	-0.38	0.51	-0.75	.46
Trust	-0.51	0.38	-1.34	.18
Ownership	-2.60	0.59	-0.44	.66
Value × Ownership	0.22	0.22	0.99	.32
Risk × Ownership	-0.22	0.49	-0.45	.65
Privacy concerns × Ownership	0.35	0.44	0.80	.42
Trust × Ownership	0.09	0.32	0.30	.77

Note. Value (M = 4.40), Risk (M = 4.17), Privacy concerns (M = 3.64), and Trust (M = 3.90) were means-centered. Ownership was effect coded with owner serving as the {-1} baseline.

5.5 Thematic analysis

The open-ended responses from the Vignette Survey were analyzed with a thematic analysis. This approach allows individual responses to be described in themes. One advantage to this approach is that it is not theory-driven, making it highly flexible in describing rich and complex data (Braun & Clark, 2006).

In the first step of the thematic analysis method, two researchers independently reviewed the responses to familiarize themselves with the data. Once they were familiar with the data, the researchers inductively generated themes to describe each response (Kurasaki, 2000). Together, the researchers then reviewed the themes and agreed on which themes to include in the codebook. Once the codebook was finalized, the two researchers independently reviewed the data again and assigned each response to a

theme (King, 2004). A Krippendorff's alpha was conducted to test the inter-rater reliability (Krippendorff, 2011; Nili, Tate, & Baros, 2017). The values of this metric range from zero to one, where zero signifies perfect disagreement and one signifies perfect agreement between the coders. A Krippendorff's alpha value of at least .67 is considered acceptable (Krippendorff, 2004); values from this analysis can be found in Tables 9–11.

Participants' survey responses (see Tables 9–11) demonstrated that when it comes to self-reported disclosure, there is a trade-off between the user value of the use case and the negative consequences associated with disclosing data. Use cases were rated as valuable when they were (1) seen as a necessity, when (2) the smart home added convenience to their lives, and when (3) the use case represented something users wanted to do frequently. For example, Participant 7 noted that *"I like to check the weather when I am heading outside so I can dress accordingly. I check the weather very frequently."*

Table 9

Themes of Value Ratings

Key themes	Agreement (%)	Krippendorff's Alpha
Convenience		
Weather	80%	0.55
Find phone	73%	0.44
Shop	87%	0.71
Make a phone call	76%	0.51
Calendar event	76%	0.44
Frequency of use		
Check the weather	76%	0.41
Find phone	77%	0.54
Shop	78%	0.52
Make a phone call	82%	0.44
View video feed	82%	0.57
Calendar event	74%	0.45
Trust		
Check the weather	93%	0.03
Find phone	96%	0.65
Shop	95%	0.74
Make a phone call	99%	0.80
View video feed	96%	0.54

Many of the responses about privacy concerns mentioned malicious data loss, where malicious actors (i.e., hackers) can get access data without explicit permission from the owner. Participant 6 said, *"All this data in the wrong hands is dangerous. People have a right to privacy. Not because they are doing something wrong but because that is a lot of data on individuals. What could they do with it? We don't even know yet. So far, buying it and selling it to other companies — which is super weird — it's our data. People should be protected."* Another participant explained their sense of helplessness on the ever-growing digital footprint of their behavior using IoT devices. Participant 50 said, *"I want my personal data and privacy to be kept private. But as I*

have learned, nothing on internet is safe and I am really insecure about my data getting leaked.

Table 10

Themes of Privacy-Concern Ratings

Key themes	Agreement (%)	Krippendorff's Alpha
Malicious data loss	70.1%	0.28
Tech limitations and weaknesses	62.9%	0.21
Trade-offs with use case value	66%	0.25
Digital footprint	68%	0.22
Trust	80.4%	0.01

Many participants discussed that people's data is often used for economic gain, whether that is by the device manufacturer (i.e., product company) or an adversary seeking to steal data. Participant 71 reported that, *"Being in the cybersecurity field, I have come to understand just how much data is being sold, leaked or harvested to be used without consumer consent or deceiving consumers into allowing companies to use their data."* Participant 98 expressed lack of trust in smart home devices as a whole by saying, *"I don't trust that the smart device would only access whatever I gave permissions for only for the app/situation I gave it permission to access whatever. I also think they could be hacked or accessed by someone else too easily."*

Table 11

Themes of Trust Ratings

Key themes	Agreement (%)	Krippendorff's Alpha
Malicious data loss	83.7%	0.65
Tech limitations and weaknesses	61.2%	0.20
User tech literacy	95.9%	0.65
Economic gain (adversary)	93.9%	0.84
Trade-offs with use case value	62.2%	0.21
Excessive requests	86.7%	0.25

Use case value, privacy concerns, and trust had overlapping themes, touching on (1) known technology limitations and weaknesses (e.g., inadequate security protocols); (2) whether users trusted the device manufacturing company; and (3) users' awareness of malicious agents that seek to steal their personal information for economic and/or social gain.

CHAPTER 6

DISCUSSION

6.1 Theoretical Implications

The findings of this study support the Privacy Paradox in the context of smart home usage (Hypothesis 1), as participants behaviorally disclosed more personal information during the Smart Home Interaction Task than they said they would be willing to disclose (c.f., Kokolakas, 2017; Awad & Krishnan, 2006; Taddicken, 2017; Acquisti & Grossklags, 2005; Cvcek et al., 2006). The correlation between the self-reported disclosure and behavior was statistically significant, but the strength of the relationship was low, consistent with existing research (e.g., Keith et al., 2013). Participants' intent to disclose their data did not reliably predict their disclosure behavior. This was likely due to the fact that overall, participants disclosed data more often than not. Over 90% of participants disclosed their personal information for each data type (location, phone number, microphone, and camera), with the exception of credit card payment information, which only 68% of respondents chose to disclose.

This research adds to the existing Privacy Calculus literature by identifying factors (use case value, privacy concerns, and trust) that participants considered when expressing their self-reported disclosure and disclosure behavior on a smart home, such as value (Zlatolas et al., 2015), privacy concerns (Gerber et al., 2018), and trust (Ackerman et al., 1999; Earp & Baumer, 2003; McKnight et al., 2011). It should be noted that the factors tested may not comprehensively describe the Privacy Calculus model. Some studies report that data sensitivity (Kehr et al., 2015), privacy literacy (Debatin et al., 2009; Park, 2013; Trepte & Dienlin, 2015), and past disclosure (Culnan

& Armstrong, 1999) are additional factors involved in data disclosure decisions. The thematic analysis in this study also identified several factors that may be predictors of disclosure decisions such as technology literacy, experience with personal data loss or abuse, and how the device behaves.

One inconsistency between these research findings and the existing literature is that risk did not, in this study, have a significant relationship with either self-reported or behavioral disclosure. Previous research has found that risk informs both self-report (e.g., Naeini et al., 2017) and behavioral disclosure (e.g., Norberg et al., 2007). This departure may be due to the way that the tasks in this study were designed, such that a clear signal on risk was not obtained. According to the Description-Experience Gap, risk perceptions are impacted by previous experience with a specific risk or by being provided with explicit description of the risk(s) of a person's behavior (Hadar & Fox; 2009; Hau et al., 2008; Wulff et al., 2015; Wulff et al., 2018). This study did not provide explicit information about the risks involved with data disclosure, instead it required participants to defer to their own experiences. However, participants might not (knowingly) have much, if any, negative experiences with sharing their personal information, so the results may have been different had there been overt information about risks of data disclosure.

Use case value, privacy concerns, and trust were confirmed predictors of self-reported disclosure. Themes derived from participants' responses in the thematic analysis included their concerns about disclosing their personal information came from the possibility that their data would be shared, without permission, to recipients outside of the smart home device itself or that this data would be acquired by external, ill-

intentioned adversaries. Another concern participants had was with respect to the limitations or weaknesses of the technology itself. If the security protocols put in place are inadequate, users' data are more vulnerable to being acquired without their explicit permission. Technical limitations were also tied to participants' sense of trust that their personal information would remain private. People were less concerned when they trusted the company that built the smart home device. They also acknowledged that adversaries aiming to steal personal data from others exist, which generally lowered their sense of trust in smart home technology overall.

Value was found to be the only predictor of disclosure behavior. It was observed that participants went quickly through the smart home interaction task, taking less time to think through the decision to disclose their data for each use case. This may indicate that people are more thoughtful about their disclosure decisions when asked explicitly if they would disclose rather than when they are actually using a device. In the vignette survey, participants responded that their ratings for use case value were explained by the practicality of using a smart home. This included the added convenience of using the smart home for a particular use case, out of necessity or need for a use case on a smart home, and the frequency with which they use a smart home for a specific use case. These reasonings hold true for both self-reported disclosure and behavior.

6.2 Practical Implications

These findings may also impact the IoT market in general. Consumers expect their data to be secure and remain private, but their behavior does not always support this notion. Tech companies can potentially gain an edge in the smart home device market by improving the user interface and improve data collection protocols to only

collect information when the device absolutely needs it. Implementing more privacy-forward user interfaces can supplement existing security protocols, which can then optimize user data protection. In doing so, users may put forth more trust in the company and its smart home devices, which may increase the adoption of those devices.

6.3 Limitations

Individual differences such as age, gender, ethnicity, and socio-economic status may have impacted these results. Most participants involved in this study were females in their early to mid-twenties, and therefore self-reported disclosure and/or behavior may not generalize to other populations. Predictors such as privacy concerns do vary across demographics. Previous research has demonstrated that women generally have more privacy concerns than men, (Sheehan, 1999; Hoy & Milne, 2010; Rowan & Dehlinger, 2014).

The Smart Home Interaction Task was designed to closely replicate what using a physical smart home would look like. However, participants interacted with a prototype—not a physical smart home device in their own home—so there may be question of true ecological validity. This may have caused them to disclose more of their personal information or even provide false information (e.g., a fake name) because the consequences of data disclosure might not have been considered as heavily as if it were done in-person on a real smart home device. Although the accuracy of the data disclosed by participants was not checked, many participants reported that they naturally provide false data in exchange for a desired use case, outcome, or reward, consistent with other studies (Steinfeld, 2015). That is, participants' disclosure behavior

may have been more liberal than it would have been in a natural environment. For example, some websites request email addresses in exchange for a discount, so people sometimes provide fake email addresses (Steinfeld, 2015). Additionally, not all scenarios provided in the Vignette Survey were realistic. For example, a smart home is unlikely to request a person's credit card information to report the local weather to the user. Future research should add a task within the study for participants to explicitly identify what use cases and data types they provided false information on. Responses containing accurate personal information would better represent people's disclosure behaviors.

Trust was left open to interpretation by participants and was not measured with a validated instrument. There are individual differences in the way that trust is interpreted in different contexts (Mai, 2020), so the granularity of trust's impact on self-reported and behavioral disclosure is not clear. In the future, researchers should explore how people define trust in the context of data disclosure with IoT devices

CHAPTER 7

CONCLUSION

Overall, this study connects the Privacy Calculus Model to the Privacy Paradox. Data shows that people's intentions to disclose their personal information is inconsistent with their actual disclosure behavior (i.e., the Privacy Paradox). Participants disclosed their data the majority of the time on both smart home devices and smartphones. Willingness to disclose, however, was much more conservative. Factors that impact people's self-reported disclosure and behaviors include use case value, perceived risk of data disclosure, general privacy concerns, and level of trust that their personal information will remain private. Existing research on the Privacy Paradox and Privacy Calculus is either context agnostic or grounded specifically in e-commerce and social networking sites. Today's technology is bigger and more powerful than websites, which is why evaluating something such as smart home devices that falls under the Internet of Things ecosystem sheds more generalizable findings to current technology people regularly use.

Future research should also delve into the granularity of each predictor. Risk may impact disclosure decisions in some capacity if people's experiences with negative consequences of data disclosure is recorded and/or vignettes included explicit descriptions of risk(s) associated with disclosing specific data in different contexts. Similar to risk, trust is multi-faceted. Trust may include, but not limited to, people's general propensity to trust, type of technology (e.g., smart home, smartphone, smart watch), and the product manufacturing company (e.g., Google, Apple, Amazon).

REFERENCES

REFERENCES

Ackerman, M. S., Cranor, L. F., & Reagle, J. (1999). Privacy in e-commerce: examining user scenarios and privacy preferences. *Proceedings of the 1st ACM conference on Electronic Commerce*. 1–8. https://doi.org/10.1145/336992.336995

Acquisti, A. (2004). Privacy in electronic commerce and the economics of immediate gratification. *Proceedings of the 5th ACM conference on Electronic Commerce.* 21–29. https://doi.org/10.1145/988772.988777

Acquisti, A., & Grossklags, J. (2004). Privacy attitudes and privacy behavior. In L. J. Camp & S. Lewis S. (Eds.), *Economics of information security. Advances in Information Security* (12, pp. 165–178). Springer, Boston, MA. https://doi.org/10.1007/1-4020-8090-5_13

Acquisti, A., & Grossklags, J. (2005). Privacy and rationality in individual decision making. *IEEE Security & Privacy*, *3*(1), 26–33. https://doi.org/10.1109/MSP.2005.22

Acquisti, A., & Grossklags, J. (2006). Privacy and rationality. In K. J. Strandburg & D. S. Raicu (Eds.) *Privacy and technologies of identity* (pp. 15–29). Springer. https://doi.org/10.1007/0-387-28222-X_2

Acquisti, A., John, L. K., & Loewenstein, G. (2013). What is privacy worth? *The Journal of Legal Studies*, *42*(2), 249–274. https://www.jstor.org/stable/10.1086/671754

Agre, P. E. (1994). Surveillance and capture: Two models of privacy. *The Information Society*, *10*(2), 101–127. https://doi.org/10.1080/01972243.1994.9960162

Apthorpe, N., Reisman, D., & Feamster, N. (2017a). A smart home is no castle: Privacy vulnerabilities of encrypted IoT traffic. *arXiv preprint arXiv:1705.06805*

Apthorpe, N., Reisman, D., Sundaresan, S., Narayanan, A., & Feamster, N. (2017b). Spying on the smart home: Privacy attacks and defenses on encrypted IoT traffic. *arXiv preprint. arXiv:1708.05044*

Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer Networks*, *54*(15), 2787–2805. https://doi.org/10.1016/j.comnet.2010.05.010

Auxier, B., Rainie, L., Anderson, M., Perrin, A., Kumar, M., & Turner, E. (2019). Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information. Retrieved from *Pew Research Center: Internet & Technology website*: https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/

Awad, N. F., & Krishnan, M. S. (2006). The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS Quarterly*, *30*(1) 13–28. https://doi.org/10.2307/25148715

Barth, S., & de Jong, M. D. T. (2017). The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior — A systematic literature review. *Telematics and Informatics*, *34*(7), 1038–1058. https://doi.org/10.1016/j.tele.2017.04.013

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., and Krivitsky, P. N. (2020). *Package lme4* (Version 1.1-25) [Computer software]. R.

Bauer, C., & Schiffinger, M. (2016). Perceived risks and benefits of online self-disclosure: affected by culture? A meta-analysis of cultural differences as moderators of privacy calculus in person-to-crowd settings. *Proceedings of the Twenty-Fourth European Conference on Information Systems (ECIS)*. https://aisel.aisnet.org/ecis2016_rp/68

Bellemare, J. (2018). What are your odds of getting your identity stolen? *Identity & Privacy: Personal*. Retrieved from https://www.identityforce.com/blog/identity-theft-odds-identity-theft-statistics.

Beresford, A. R., Kübler, D., & Preibusch, S. (2012). Unwillingness to pay for privacy: A field experiment. *Economics Letters, 117*(1), 25–27. https://doi.org/10.1016/j.econlet.2012.04.077

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Brown, B. (2001). Studying the Internet experience. HP laboratories technical report HPL, 49.

Camilleri, A. R., & Newell, B. R. (2013). Mind the gap? Description, experience, and the continuum of uncertainty in risky choice. *Progress in Brain Research*, *202*, 55–71. https://doi.org/10.1016/B978-0-444-62604-2.00004-6

Campbell, J. E., & Carlson, M. (2002). Panopticon.com: Online surveillance and the commodification of privacy. *Journal of Broadcasting & Electronic Media*, *46*(4), 586–606. https://doi.org/10.1207/s15506878jobem4604_6

Carrascal, J. P., Riederer, C., Erramilli, V., Cherubini, M., & de Oliveira, R. (2013). Your browsing behavior for a big mac: Economics of personal information online. *Proceedings of the 22nd international conference on World Wide Web,* 189–200. https://doi.org/10.1145/2488388.2488406

Castelfranchi, C., & Falcone, R. (2005). Socio-cognitive theory of trust. *J. Pitt. London: Wiley.*

Chan, H., & Perrig, A. (2003). Security and privacy in sensor networks. *Computer*, *36*(10), 103–105. https://doi.org/10.1109/MC.2003.1236475

Chen, J., Zhang, Y., & Heath, R. (2001). An Exploratory Investigation of the Relationships between Consumer Characteristics and Information Privacy. *Marketing Management Journal*, *11*(1), 73–81. https://doi.org/10.1108/13522750610640558

Christofides, E., Muise, A., & Desmarais, S. (2009). Information disclosure and control on Facebook: Are they sides of the same coin or two different processes? *CyberPsychology & Behavior, 12*(3), 341–345. https://doi.org/10.1089/cpb.2008.0226

Chung, H., Iorga, M., Voas, J., & Lee, S. (2017). Alexa, can I trust you? *Computer, 50*(9), 100–104. https://doi.org/10.1109/MC.2017.3571053

Cohen, J. (1992). A power primer. *Quantitative Methods in Psychology, 112(1)*, 155–159. https://doi.org/10.1037//0033-2909.112.1.155

Cowan, B. R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., Bandeira, N. (2017). What can I help you with?: infrequent users' experiences of intelligent personal assistants. *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services. 1–12*. https://doi.org/10.1145/3098279.3098539

Crager, K., & Maiti, A. (2017). Information leakage through mobile motion sensors: User awareness and concerns. [Paper presentation] *The Proceedings of the European Workshop on Usable Security (EuroUSEC).* https://doi.org/10.14722/EUROUSEC.2017.23013

Culnan, M. J., & Armstrong, P. K. (1999). Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization Science, 10*(1), 104–115. https://doi.org/10.1287/orsc.10.1.104

Cvrcek, D., Kumpost, M., Matyas, V., & Danezis, G. (2006). A study on the value of location privacy. *Proceedings of the 5th ACM workshop on Privacy in electronic society*, 109–118. https://doi.org/10.1145/1179601.1179621

Davis, F. D. (1985). A technology acceptance model for empirically testing new end-user information systems: Theory and results. *Massachusetts Institute of Technology.*

Davis, F. D., & Venkatesh, V. (1996). A critical assessment of potential measurement biases in the technology acceptance model: three experiments. *International*

Journal of Human-Computer Studies, 45(1), 19–45. https://doi.org/10.1006/ijhc.1996.0040

Debatin, B., Lovejoy, J. P., Horn, A.-K., & Hughes, B. N. (2009). Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of Computer-Mediated Communication*, *15*(1), 83–108. https://doi.org/10.1111/j.1083-6101.2009.01494.x

Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research, 17*(1), 61–80. https://doi.org/10.1287/isre.1060.0080

Earp, J. B., & Baumer, D. (2003). Innovative web use to learn about consumer behavior and online privacy. *Communications of the ACM, 46*(4), 81–83. https://doi.org/10.1145/641205.641209

Egelman, S., Felt, A. P., & Wagner, D. (2013). Choice architecture and smartphone privacy: There's a price for that. *The Economics of Information Security and Privacy, 17*(1), 211–236. https://doi.org/10.1007/978-3-642-39498-0_10

Equifax. (2017). Statement of record regarding the extent of the cybersecurity incident announced on September 7, 2017 [Press release]. Retrieved from https://www.sec.gov/Archives/edgar/data/33185/000119312518154706/d583804dex991.htm

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*(443). https://doi.org/10.3389/fnhum.2014.00443

Gelman, A., and Hill, J. (2006). Multilevel linear models: the basics. *Data analysis using regression and multilevel/hierarchical models (251–278)*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790942.016

Gerber, N., Gerber, P., & Volkamer, M. (2018). Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. *Computers & Security*, 77, 226–261. https://doi.org/10.1016/j.cose.2018.04.002

Govani, T., & Pashley, H. (2005). Student awareness of the privacy implications when using Facebook. [Unpublished manuscript]. *Department of Mathematical Science, Carnegie Mellon University,* 9, 1–17.

Hadar, L., & Fox, C. R. (2009). Information asymmetry in decision from description versus decision from experience. *Judgment and Decision Making*, *4*(4), 317–325.

Hann, I.H., Hui, K.L., Lee, T., & Png, I. (2002). Online information privacy: Measuring the cost-benefit trade-off. *Proceedings of the ICIS*. 1.

Hann, I.H., Hui, K.L., Lee, S.Y. T., & Png, I.P. (2007). Overcoming online information privacy concerns: An information-processing theory approach. *Journal of Management Information Systems*, *24*(2), 13–42. https://doi.org/10.2753/MIS0742-1222240202

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description–experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making, 21*(5), 493–518. https://doi.org/10.1002/bdm.598

Hauswald, J., Laurenzano, M. A., Zhang, Y., Li, C., Rovinski, A., Khurana, A., Tang, L. (2015). Sirius: An open end-to-end voice and vision personal assistant and its implications for future warehouse scale computers. *Proceedings of the Twentieth International Conference on Architectural Support for Programming Languages and Operating Systems. 223–238*. https://doi.org/10.1145/2694344.2694347

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15*(8), 534–539. https://doi.org/10.1111/j.0956-7976.2004.00715.x

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences*, *13*(12), 517–523. https://doi.org/10.1016/j.tics.2009.09.004

Hoffman, D. L., Novak, T. P., & Peralta, M. (1999). Building consumer trust online. *Communications of the ACM, 42*(4), 80–85. https://doi.org/10.1145/299157.299175

Hoy, M. G., & Milne, G. (2010). Gender differences in privacy-related measures for young adult Facebook users. *Journal of Interactive Advertising*, *10*(2), 28-45. https://doi.org/10.1080/15252019.2010.10722168

Huberman, B. A., Adar, E., & Fine, L. R. (2005). Valuating privacy. *IEEE Security & Privacy, 3*(5), 22–25. https://doi.org/10.1109/MSP.2005.137

Hughes-Roberts, T. (2013). Privacy and social networks: Is concern a valid indicator of intention and behaviour? Proceedings of the *2013 International Conference on Social Computing.* 909–912. https://doi/org/10.1109/SocialCom.2013.140

Jacobsson, A., Boldt, M., & Carlsson, B. (2016). A risk analysis of a smart home automation system. *Future Generation Computer Systems*, *56*, 719–733. https://doi.org/10.1016/j.future.2015.09.003

Jennings, N. R. (2000). On agent-based software engineering. *Artificial Intelligence*, *117*(2), 277–296. https://doi.org/10.1016/S0004-3702(99)00107-1

Joinson, A. N., Houghton, D. J., Vasalou, A., & Marder, B. L. (2011). Digital crowding: Privacy, self-disclosure, and technology. *Privacy Online,* 33–45. https://doi.org/10.1007/978-3-642-21521-6_4

Jöst, M., Häußler, J., Merdes, M., & Malaka, R. (2005). Multimodal interaction for pedestrians: an evaluation study. *Proceedings of the 10th International Conference on Intelligent User Interfaces.* 59–66. https://doi.org/10.1145/1040830.1040852

Kang, C., & Frenkel, S. (2018). Facebook says Cambridge Analytica harvested data of up to 87 million users. *New York Times*, 4.

Kehr, F., Kowatsch, T., Wentzel, D., & Fleisch, E. (2015). Blissfully ignorant: the effects of general privacy concerns, general institutional trust, and affect in the privacy calculus. *Information Systems Journal*, *25*(6), 607–635. https://doi.org/10.1111/isj.12062

Keith, M. J., Thompson, S. C., Hale, J., Lowry, P. B., & Greer, C. (2013). Information disclosure on mobile devices: Re-examining privacy calculus with actual user behavior. *International Journal of Human-Computer Studies*, *71*(12), 1163–1173. https://doi.org/10.1016/j.ijhcs.2013.08.016

Kerry, C. (2018). Why protecting privacy is a losing game today--and how to change the game. *Brookings*. Retrieved from Brookings website: https://www.brookings.edu/research/why-protecting-privacy-is-a-losing-game-today-and-how-to-change-the-game/.

Khan, R., Khan, S. U., Zaheer, R., & Khan, S. (2012). Future internet: The internet of things architecture, possible applications and key challenges. *Proceedings of the 10th International Conference on Frontiers of Information Technology (FIT), Institute of Electrical and Electronics Engineers Inc*. https://doi/org/10.1109/FIT.2012.53

King, N. (2004). Using Templates in the Thematic Analysis of Text. *Essential Guide to Qualitative Methods in Organizational Research*, 257–260. https://doi.org/0.4135/9781446280119.n21

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, 64, 122–134. https://doi.org/10.1016/j.cose.2015.07.002

Krasnova, H., Günther, O., Spiekermann, S., & Koroleva, K. (2009). Privacy concerns and identity in online social networks. *Identity in the Information Society*, *2*(1), 39–63.

Krasnova, H., Veltri, N. F., & Günther, O. (2012). Self-disclosure and privacy calculus on social networking sites: The role of culture. *Business & Information Systems Engineering*, *4*(3), 127–135. https://doi.org10.1007/s12599-012-0216-6

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology. Second Edition*. Thousand Oaks, CA: Sage.

Krippendorff, K. (2011). Computing Krippendorff's alpha-reliability. Retrieved from http://repository.upenn.edu/asc_papers/43.

Kurasaki, K. S. (2000). Intercoder reliability for validating conclusions drawn from open-ended interview data. *Field Methods*, *12*(3), 179–194. https://doi.org/10.1177/1525822X0001200301

Lankton, N., McKnight, D. H., & Thatcher, J. B. (2013). Incorporating trust-in-technology into Expectation Disconfirmation Theory. *The Journal of Strategic Information Systems, 23*(2). https://doi.org/10.1016/j.jsis.2013.09.001

Laufer, R. S., & Wolfe, M. (1977). Privacy as a concept and a social issue: A multidimensional developmental theory. *Journal of Social Issues*, *33*(3), 22–42. https://doi.org/10.1111/j.1540-4560.1977.tb01880.x

Lenth, R., Singmann, H., Love, J., Buerkner, P., and Herve, M. (2019). *Package emmeans* (Version 1.3.4) [Computer software]. R.

Loyd, B., Kennedy, C., & Yoder, P. (2013). Quantifying contingent relations from direct observation data: Transitional probability comparisons versus Yule's Q. *Journal of Applied Behavior Analysis*, *46*(2), 479–497. https://doi.org/10.1002/jaba.45

Mai, J. E. (2016). Big data privacy: The datafication of personal information. *The Information Society*, *32*(3), 192–199. https://doi.org/10.1080/01972243.2016.1153010

Mai, J. E. (2020). Three models of privacy: New perspectives on informational privacy. *Nordicom Review*, *37*(1), 171–175. https://doi.org/10.1515/nor-2016-0031

Malheiros, M., Preibusch, S., & Sasse, M. A. (2013). “Fairly truthful”: The impact of perceived effort, fairness, relevance, and sensitivity on personal data disclosure. *Proceedings of the International Conference on Trust and Trustworthy Computing,* 250–266. https://doi.org/10.1007/978-3-642-38908-5_19

Martin, K. (2018). The penalty for privacy violations: How privacy violations impact trust online. *Journal of Business Research*, *82*, 103–116. https://doi.org/10.1016/j.jbusres.2017.08.034

Mayr, S., Erdfelder, E., Buchner, A., & Faul, F. (2007). A short tutorial of GPower. *Tutorials in Quantitative Methods for Psychology, 3(2)*, 51–59. https://doi/org/10.20982/tqmp.03.2.p051

Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems (TMIS)*, *2*(2), 12. https://doi.org/10.1145/1985347.1985353

Metzger, M. J. (2004). Privacy, Trust, and Disclosure: Exploring Barriers to Electronic Commerce. *Journal of Computer-Mediated Communication, 9*(4). https://doi.org/10.1111/j.1083-6101.2004.tb00292.x

Mihale-Wilson, C., Zibuschka, J., & Hinz, O. (2017). About user preferences and willingness to pay for a secure and privacy protective ubiquitous personal assistant. *Proceedings of the 25th European Conference on Information Systems (ECIS).*

Morando, F., Iemma, R., & Raiteri, E. (2014). Privacy evaluation: What empirical research on users' valuation of personal data tells us. *Internet Policy Review*, *3*(2). https://doi.org/10.14763/2014.2.283

Mothersbaugh, D. L., Foxx, W. K., Beatty, S. E., & Wang, S. (2012). Disclosure antecedents in an online service context: The role of sensitivity of information. *Journal of Service Research*, *15*(1), 76–98. https://doi.org/10.1177/1094670511424924

Newell, B. R., & Rakow, T. (2007). The role of experience in decisions from description. *Psychonomic Bulletin & Review*, *14*(6), 1133–1139. https://doi.org/10.3758/BF03193102

Nili, A., Tate, M., & Barros, A. (2017). A critical analysis of inter-coder reliability methods in information systems research. *Proceedings of the 28th Australasian Conference on Information Systems*.

Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*: Stanford University Press.

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information self-reported disclosure versus behaviors. *Journal of Consumer Affairs*, *41*(1), 100–126. https://doi.org/10.1111/j.1745-6606.2006.00070.x

Nov, O., & Wattal, S. (2009). Social computing privacy concerns: antecedents and effects. *Proceedings of the SIGCHI conference on human factors in computing systems*. https://doi/org/10.1145/1518701.1518754

Park, Y. J. (2013). Digital literacy and privacy behavior online. *Communication Research*, *40*(2), 215–236. https://doi.org/10.1177/0093650211418338

Rowan, M., & Dehlinger, J. (2014). Observed gender differences in privacy concerns and behaviors of mobile device end users. *Procedia Computer Science*, *37*, 340–347. https://doi.org/10.1016/j.procs.2014.08.050

R Core Team. (2020). *R: A language and environment for statistical computing* (Version 4.0.3) [Computer software]. R Foundation for Statistical Computing.

Rubin, Z. (1975). Disclosing oneself to a stranger: Reciprocity and its limits. *Journal of Experimental Social Psychology, 11*(3), 233–260. https://doi.org/10.1016/S0022-1031(75)80025-4

Sayre, S., & Horne, D. (2000). Trading secrets for savings: How concerned are consumers about club cards as a privacy threat? *Advances in Consumer Research*, *27*, 151–155.

Schwartz, E., & Srail, E. (2014). *Prototyping Essentials with Axure*. Packt Publishing Ltd.

Sheehan, K. B. (1999). An investigation of gender differences in on-line privacy concerns and resultant behaviors. *Journal of Interactive Marketing*, *13*(4), 24–38. https://doi.org/10.1002/(SICI)1520-6653(199923)13:4<24::AID-DIR3>3.0.CO;2-O

Schomakers, E.-M., Lidynia, C., Müllmann, D., & Ziefle, M. (2019). Internet users' perceptions of information sensitivity–insights from Germany. *International Journal of Information Management, 46,* 142–150. https://doi.org/10.1016/j.ijinfomgt.2018.11.018

Smith, S. (2018). *Digital voice assistants in use to triple to 8 billion by 2023, driven by smart home devices*. Juniper Research. Retrieved from https://www.juniperresearch.com/press/press-releases/digital-voice-assistants-in-use-to-8-million-2023.

Spiekermann, S., Grossklags, J., & Berendt, B. (2001). E-privacy in 2nd generation E-commerce: privacy preferences versus actual behavior. *Proceedings of the 3rd ACM Conference on Electronic Commerce.* 38–47. https://doi.org/10.1145/501158.501163

Steinfeld, N. (2015). Trading with privacy: The price of personal information. *Online Information Review, 39*(7), 923–938. https://doi.org/10.1108/OIR-05-2015-0168

Taddicken, M. (2014). The 'privacy paradox' in the social web: The impact of privacy concerns, individual characteristics, and the perceived social relevance on different forms of self-disclosure. *Journal of Computer-Mediated Communication, 19*(2), 248–273. https://doi.org/10.1111/jcc4.12052

Tait, S. E., & Jeske, D. (2015). Hello Stranger!: Trust and Self-Disclosure Effects on Online Information Sharing. *International Journal of Cyber Behavior, Psychology and Learning (IJCBPL), 5*(1), 42–55. https://doi.org/10.4018/ijcbpl.2015010104

Trepte, S., Teutsch, D., Masur, P. K., Eicher, C., Fischer, M., Hennhöfer, A., & Lind, F. (2015). Do People Know About Privacy and Data Protection Strategies? Towards the "Online Privacy Literacy Scale" (OPLIS). In S. Gutwirth, R. Leenes, & P. de

Hert (Eds.), *Reforming European Data Protection Law* (pp. 333–365). Dordrecht: Springer Netherlands.

Tufekci, Z. (2008). Can you see me now? Audience and disclosure regulation in online social network sites. *Bulletin of Science, Technology & Society, 28*(1), 20–36. https://doi.org/10.1177/0270467607311484

Ur, B., Leon, P. G., Cranor, L. F., Shay, R., & Wang, Y. (2012). Smart, useful, scary, creepy: perceptions of online behavioral advertising. *Proceedings of the 8th Symposium on Usable Privacy and Security (SOUPS).* 1–15. https://doi.org/10.1145/2335356.2335362

Vroom, V. H. (1964). *Work and motivation*. New York: Wiley.

Weisbaum. (2018). Trust in Facebook has dropped by 66 percent since the Cambridge Analytica scandal. *NBC News*. Retrieved from https://www.nbcnews.com/business/consumer/trust-facebook-has-dropped-51-percent-cambridge-analytica-scandal-n867011.

Wortmann, F., & Flüchter, K. (2015). Internet of Things: Technology and Value Added. *Business & Information Systems Engineering, 57*(3), 221–224. https://doi.org/10.1007/s12599-015-0383-3

Wu, K.W., Huang, S.Y., Yen, D.C., & Popova, I. (2012). The effect of online privacy policy on consumer privacy concern and trust. *Computers in Human Behavior, 28*(3), 889–897. https://doi.org/10.1016/j.chb.2011.12.008

Wulff, D. U., Hills, T. T., & Hertwig, R. (2015). Online product reviews and the description–experience gap. *Journal of Behavioral Decision Making*, *28*(3), 214–223. https://doi.org/10.1002/bdm.1841

Wulff, D.U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, *144*(2), 140–176. https://doi.org/10.1037/bul0000115

Young, A. L., & Quan-Haase, A. (2009). Information revelation and internet privacy concerns on social network sites: a case study of Facebook. *Proceedings of the 4th International Conference on Communities and Technologies.* https://doi.org/10.1145/1556460.1556499

Zibuschka, J., Nofer, M., Zimmermann, C., & Hinz, O. (2019). Users' preferences concerning privacy properties of assistant systems on the internet of things. *Proceedings of the 25th Americas Conference on Information Systems*.

Zlatolas, L. N., Welzer, T., Heričko, M., & Hölbl, M. (2015). Privacy antecedents for SNS self-disclosure: The case of Facebook. *Computers in Human Behavior*, 45, 158–167. https://doi.org/10.1016/j.chb.2014.12.012

APPENDICES

APPENDIX A

VIGNETTE SURVEY

**To control for order effects, the blocks were randomized, as well as each data type within each use case block.*

Check the weather

- A device requests your location data to check the weather. Will you provide the requested information?
- A device requests your phone number to check the weather. Will you provide the requested information?
- A device requests your credit card number to check the weather. Will you provide the requested information?
- A device requests access to your microphone to check the weather. Will you provide the requested information?
- A device requests access to your camera to check the weather. Will you provide the requested information?
- A device requests access to your calendar to check the weather. Will you provide the requested information?
- A device requests access to your photos to check the weather. Will you provide the requested information?
- A device requests access to your account on a third-party app to check the weather. Will you provide the requested information?

Find phone

- You misplaced your phone. Will you provide your location to locate it?
- You misplaced your phone. Will you provide your phone number to locate it?
- You misplaced your phone. Will you provide your credit card number to locate it?
- You misplaced your phone. Will you grant access to your microphone to locate it?
- You misplaced your phone. Will you grant access to your camera to locate it?
- You misplaced your phone. Will you grant access to your calendar to locate it?
- You misplaced your phone. Will you grant access to your photos to locate it?
- You misplaced your phone. Will you grant access to your account on a third-party app to locate it?

Shop online

- You are online shopping on a device. Will you provide your location data to do so?
- You are online shopping on a device. Will you provide your phone number to do so?
- You are online shopping on a device. Will you provide your credit card number to do so?
- You are online shopping on a device. Will you grant access to your microphone to do so?
- You are online shopping on a device. Will you grant access to your camera to do so?

- You are online shopping on a device. Will you grant access to your calendar to do so?
- You are online shopping on a device. Will you grant access to your photos to do so?
- You are online shopping on a device. Will you grant access to your account on a third-party app to do so?

Make a phone/audio call

- A device requests your location data to make a phone/audio call. Will you provide this information?
- A device requests your phone number to make a phone/audio call. Will you provide this information?
- A device requests your credit card number to make a phone/audio call. Will you provide this information?
- A device requests access to your microphone to make a phone/audio call. Will you provide this information?
- A device requests access to your camera to make a phone/audio call. Will you provide this information?
- A device requests access to your calendar to make a phone/audio call. Will you provide this information?
- A device requests access to your photos to make a phone/audio call. Will you provide this information?

- A device requests access to your account on a third-party app to make a phone/audio call. Will you provide this information?

View live video feed

- A device requests your location data to view a live video feed of a room in your house. Will you provide this information?
- A device requests your phone number to view a live video feed of a room in your house. Will you provide this information?
- A device requests your credit card number to view a live video feed of a room in your house. Will you provide this information?
- A device requests access to your microphone to view a live video feed of a room in your house. Will you provide this information?
- A device requests access to your camera to view a live video feed of a room in your house. Will you provide this information?
- A device requests access to your calendar to view a live video feed of a room in your house. Will you provide this information?
- A device requests access to your photos to view a live video feed of a room in your house. Will you provide this information?
- A device requests access to your account on a third-party app to view a live video feed of a room in your house. Will you provide this information?

Set a calendar event

- A device requests your location data to set a calendar event. Will you provide this information?

- A device requests your phone number to set a calendar event. Will you provide this information?
- A device requests your credit card number to set a calendar event. Will you provide this information?
- A device requests access to your microphone to set a calendar event. Will you provide this information?
- A device requests access to your camera to set a calendar event. Will you provide this information?
- A device requests access to your calendar to set a calendar event. Will you provide this information?
- A device requests access to your photos to set a calendar event. Will you provide this information?
- A device requests access to your account on a third-party app to set a calendar event. Will you provide this information?

Stream media

- A device requests your location data to stream media (e.g., music, videos). Will you provide this information?
- A device requests your phone number to stream media (e.g., music, videos). Will you provide this information?
- A device requests your credit card number to stream media (e.g., music, videos). Will you provide this information?

- A device requests access to your microphone to stream media (e.g., music, videos). Will you provide this information?
- A device requests access to your camera to stream media (e.g., music, videos). Will you provide this information?
- A device requests access to your calendar to stream media (e.g., music, videos). Will you provide this information?
- A device requests access to your photos to stream media (e.g., music, videos). Will you provide this information?
- A device requests access to your account on a third-party app to stream media (e.g., music, videos). Will you provide this information?

Follow-up questions to each vignette

- How valuable is being able to [USE CASE] to you?

  Not at all valuable [1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7] Extremely valuable
- How risky is it to you to disclose your [DATA TYPE]?

  Not at all risky [1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7] Extremely risky
- How concerned are you about disclosing your [DATA TYPE]?

  Not at all sensitive [1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7] Extremely sensitive
- How trustworthy do you feel devices are in keeping your [DATA TYPE] private?

  Not at all trustworthy [1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7] Extremely trustworthy

High-level ratings (end of survey)

- How trustworthy do you feel devices are in keeping your data private?

Not at all trustworthy [1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7] Extremely trustworthy

- Explain why you rated trustworthiness the way you did.

  Open-ended response

- How valuable is it to you to be able to [USE CASE] on a device?

  Not at all valuable [1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7] Extremely valuable

- Explain why you rated the value for [USE CASE].

  Open-ended response

- Regarding your data privacy, how risky do you feel it is to disclose your [DATA TYPE] with a device?

  Not at all risky [1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7] Extremely risky

- In general, how concerned are you about your data privacy?

  Not at all concerned [1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7] Extremely concerned

APPENDIX B

MULTI-LEVEL LOGISTIC REGRESSION CODE

```
library(tidyverse)
library(lme4)
library(emmeans)
library(effects)
library(ggplot2)
library(car)

SR <- read.csv("SRrestrx.csv")[,2:13]
behavior <- read.csv("CarterData_behav.csv")[,2:11]

SR$ownership <- as.factor(SR$ownership)

contrasts(SR$dataType) <- contr.sum(8)
contrasts(SR$useCase) <- contr.sum(7)
contrasts(SR$ownership) <- contr.sum(2)

SR$value.c <- scale(SR$value, scale = F, center = T)
SR$risk.c <- scale(SR$risk, scale = F, center = T)
SR$trust.c <- scale(SR$trust, scale = F, center = T)
SR$concern.c <- scale(SR$conc, scale = F, center = T)


lm1 <- glmer(discloseChoice ~ value.c + risk.c + concern.c + trust.c + (1|id) + (1|us
eCase) + (1|dataType), data = SR, family = "binomial")

summary(lm1)
vif(lm1)

data <- data.frame(Effect("value.c", lm1, xlevels = list(value.c = c(1, 3, 5, 7))))

data$LSE <- data$fit - data$se
data$USE <- data$fit + data$se

ggplot(data, aes(x = value.c , y = fit, ymin = LSE, ymax = USE)) + geom_line() + geom
_ribbon(alpha = 0.5) + theme_classic(base_size = 14) + labs(y = "P(Disclosure)", x =
"Value of Use Case") + scale_y_continuous(limits = c(0, 1), labels = scales::number_f
ormat(accuracy = 0.1)) + scale_x_continuous(breaks = c(1, 2, 3, 4, 5, 6, 7))

data <- data.frame(Effect("concern.c", lm1, xlevels = list(concern.c = c(1, 3, 5, 7)) ))

data$LSE <- data$fit - data$se
data$USE <- data$fit + data$se
```

```
ggplot(data, aes(x = concern.c , y = fit, ymin = LSE, ymax = USE)) + geom_line() + ge
om_ribbon(alpha = 0.5) + theme_classic(base_size = 14) + labs(y = "P(Disclosure)", x =
"Concern Associated with Disclosure") + scale_y_continuous(limits = c(0, 1), labels =
scales::number_format(accuracy = 0.1)) + scale_x_continuous(breaks = c(1, 2, 3, 4, 5,
6, 7))

data <- data.frame(Effect("trust.c", lm1, xlevels = list(trust.c = c(1, 3, 5, 7))))

data$LSE <- data$fit - data$se
data$USE <- data$fit + data$se

ggplot(data, aes(x = trust.c , y = fit, ymin = LSE, ymax = USE)) + geom_line() + geom
_ribbon(alpha = 0.5) + theme_classic(base_size = 14) + labs(y = "P(Disclosure)", x =
"Trust") + scale_y_continuous(limits = c(0, 1), labels = scales::number_format(accura cy
= 0.1)) + scale_x_continuous(breaks = c(1, 2, 3, 4, 5, 6, 7)



lm2 <- glmer(discloseChoice ~ value.c*ownership + risk.c*ownership +
concern.c*owners hip + trust.c*ownership + (1|id) + (1|useCase) + (1|dataType), data =
SR, family = "b inomial")

summary(lm2)
vif(lm2)

emmeans(lm2, pairwise ~ ownership)$contrasts
emtrends(lm2, pairwise ~ ownership, var = "trust.c")$contrasts
emtrends(lm2, pairwise ~ ownership, var = "value.c")$contrast
emtrends(lm2, pairwise ~ ownership, var = "risk.c")$contrasts
emtrends(lm2, pairwise ~ ownership, var = "concern.c")$contrasts


$dataType <- as.factor(behavior$dataType)
behavior$ownership <- as.factor(behavior$ownership)
contrasts(behavior$dataType) <- contr.sum(5)
contrasts(behavior$useCase) <- contr.sum(7)
contrasts(behavior$ownership) <- contr.sum(2)

behavior$value.c <- scale(behavior$value, scale = F, center = T)
behavior$risk.c <- scale(behavior$risk, scale = F, center = T)
behavior$trust.c <- scale(behavior$trust, scale = F, center = T)
behavior$conc.c <- scale(behavior$conc, scale = F, center = T)
```

```
lm3 <- glmer(disclosure ~ value.c + risk.c + concern.c + trust.c + (1|id) + (1|useCas e),
data = behavior, family = "binomial"

summary(lm3)
vif(lm3)

data <- data.frame(Effect("value.c", lm3, xlevels = list(value.c = c(1, 3, 5, 7))))

data$LSE <- data$fit – data$se
data$USE <- data$fit + data$se

ggplot(data, aes(x = value.c , y = fit, ymin = LSE, ymax = USE)) + geom_line() + geom
_ribbon(alpha = 0.5) + theme_classic(base_size = 14) + labs(y = "P(Disclosure)", x =
"Use Case Value") + scale_y_continuous(limits = c(0, 1.2), labels = scales::number_fo
rmat(accuracy = 0.1)) + scale_x_continuous(breaks = c(1, 2, 3, 4, 5, 6, 7))



lm4 <- glmer(disclosure ~ value.c*ownership + risk.c*ownership + concern.c*ownership
+ trust.c*ownership + (1|id) + (1|useCase), data = behavior, family = "binomial")

summary(lm4)
vif(lm4)

emmeans(lm4, pairwise ~ ownership)$contrasts
emtrends(lm4, pairwise ~ ownership, var = "trust.c")$contrasts
emtrends(lm4, pairwise ~ ownership, var = "value.c")$contrasts
emtrends(lm4, pairwise ~ ownership, var = "risk.c")$contrasts
emtrends(lm4, pairwise ~ ownership, var = "concern.c")$contrasts



behavior$discloseChoice <- as.factor(behavior$discloseChoice) behavior$disclosure <-
as.factor(behavior$disclosure)

contrasts(behavior$discloseChoice) <- contr.sum(2)

lm6 <- glmer(disclosure ~ discloseChoice + (1|id) + (1|useCase), data = behavior, fam
ily = "binomial")

summary(lm6)
vif(lm6)
```