The Dissertation Committee for Junjie Qian Certifies that this is the approved version of the following Dissertation:

ESSAYS ON DIGITAL AND PHYSICAL CHANNELS AND MITIGATION MECHANISMS OF IDENTITY FRAUDS

Committee:

Huseyin Tanriverdi, Supervisor

Ashish Agarwal

Sirkka L Jarvenpaa

Suzanne Barber

ESSAYS ON DIGITAL AND PHYSICAL CHANNELS AND MITIGATION MECHANISMS OF IDENTITY FRAUDS

by

Junjie Qian

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2021

Abstract

ESSAYS ON DIGITAL AND PHYSICAL CHANNELS AND MITIGATION MECHANISMS OF IDENTITY FRAUDS

Junjie Qian, PhD

The University of Texas at Austin, 2021

Supervisor: Huseyin Tanriverdi

This dissertation includes two essays on identity theft. we developed a modified routine activity theory (RAT) in the first essay to explain the channels of identity frauds occurrence. This study aims to understand the antecedents of identity theft victimization at the individual level, which is important for the law enforcement agency to understand how identity theft occur and design effective deterrence measures. Different from existing studies that examine the antecedents of identity theft using the online proximity approach, we propose the novel concept of recognitive proximity. Recognitive proximity measures how closely a criminal impersonates a victim's identity by exploiting personal information in four identity dimensions, including symbolic representation, personal identity, role identity and social identity. We hypothesize that online proximity and physical proximity created by individual and organizational leakage of personal information in both digital and physical channels may result in different levels of recognitive proximity, thus

generating distinct influence on the risk of identity fraud victimizations. We test our hypotheses using a representative sample of 12,376 U.S. citizens. Our hypotheses are supported by the empirical results. In the second essay, we develop a theory to explain whether, how, and when individuals can protect themselves against the heightened identity theft (IDT) risks following a data breach. We conceptualize IDT as a multi-stage process where criminals first unlawfully obtain a person's identifying information (PII) through a data breach, then misuse the PII to assume the identity of the person, and ultimately, imposter as the person to commit the IDT crime. We distinguish if the person's PII leaks to criminals through a personal data breach or an organizational data breach. We hypothesize that preventive protections can reduce the PII leakage through personal data breaches but they can increase the PII leakage through organizational data breaches. We further hypothesize that detective protections can mitigate the heightened IDT risks following a data breach. Finally, we hypothesize that proactive protections are likely to be more effective than reactive protections in reducing IDT. We find support for these ideas in a representative sample of 66,224 citizens in the U.S.A.

Table of Contents

List of Tables

List of Figures

BEYOND PII LEAKAGE: DIGITAL AND PHYSICAL CHANNELS OF IDENTITY FRAUDS

Abstract

We extend the routine activity theory (RAT) by studying the novel concept of recognitive proximity. We theorize that individual's personal information falls into four identity dimensions, including symbolic representation, personal identity, role identity and social identity. We argue that identity theft involves the impersonation of a victim's identity in these four identity dimensions. The scope of personal information and identity dimensions exploited for impersonation determines the level of recognitive proximity between an identity thief and a victim's identity. Building on this theoretical framework, we distinguish the level of recognitive proximity created by three channels that can leak personal information, including individual online shopping activity, individual physical theft, and organizational data breach. Using a representative sample of 12,376 U.S. citizens observed in 2012, 2014, and 2016, we find that individual online shopping activity and individual physical theft increase the probability of victimization that require low to medium recognitive proximity, including credit and debit card fraud, other existing account fraud and new account fraud. We also find that organizational data breach increases the probability of all types of identity theft victimization. Our results support our hypotheses that digital and physical mechanisms of identity theft through individual and organizational channels of personal information leakage create distinct levels of recognitive proximity, thus resulting in distinct impact on different types of identity theft victimization.

Chapter 1: Introduction

Identity theft involves the misuse of personal information of other people to commit fraud or other crimes. Recent studies estimate that about 14 to 26 million people were victims of identity theft crimes in the U.S. (Harrell 2019; Marchini and Pascual 2019). That is, every 1 in 10 to 15 people was a victim of identity theft from 2016 to 2019. The total economic loss of identity theft amounts to $17.5 billion, which exceeded the combined total economic loss of $16.4 million due to all traditional property crimes (e.g., burglary, car theft, larceny theft) in 2018 (FBI 2019). During the recent pandemic, the financial loss due to identity theft has skyrocketed. In 2020, the Department of Labor's Office of the Inspector General (OIG) estimated that the U.S. lost more than $36 billion in unemployment benefits due to improper payments mainly caused by identity theft (OIG 2020). In addition to financial loss, reputation of victims of identity theft is damaged and it can take more than six months to clean the fraudulent records and recover one's good name. Many victims of identity theft also experience severe emotional distress (Harrell 2019). To help minimize identity theft risks, it is important to better understand the antecedents and occurrence mechanisms of identity theft.

IS researchers have studied privacy and security of personal information extensively (Acquisti 2004). Identity theft is closely related to this research topic as this type of crimes is a result of misuse of someone else's personal information. To better understand identity theft, we recognize that there is a large scope of personal information that falls into four different identity dimensions. Identity is the multifaceted and complex answer to the question "Who are you?". It is consisted of 1) symbolic representation, 2) personal identity, 3) role identity and 4) social identity (Camp 2004; Craig et al. 2019). For an identity thief to impersonate a victim's identity, he/she may need to exploit personal information in every identity dimension. Studies that do not consider the multi-dimensional

nature of identity implicitly assume that gaining access to individual's personally identifiable information (PII) in the symbolic representation identity dimension is sufficient for identity theft. This assumption is not necessarily valid, however, because identity theft can require a wider scope of personal information that involve every identity dimension. Thus, we examine the differences in the scope of personal information and identity dimensions involved in distinct types of fraud and identity theft. We also examine the occurrence mechanisms of each type of fraud and identity theft.

Previous identity theft research has built on routine activity theory (RAT) to explain the occurrence of identity theft. Existing RAT views the increase in identity theft victimization as a result of the shift of individual routine activities into the digital space (Eck and Clarke 2003a). Individual online activities, such as online shopping, have increased online proximity, i.e., the ease of availability of individual PII to potential offenders through the shared network (Reyns and Henson 2016). We argue that online proximity is not necessary nor sufficient for identity theft to occur. Identity theft has existed before the Internet was invented. For example, Frank Abagnale was a well-known con man who conducted identity theft cases in the 1970s. Even today personal information is contained in many physical documents. Physical proximity to these documents is also related to identity theft. Different from existing RAT, we propose the novel concept of recognitive proximity, which measures the closeness that an identity thief impersonate the victim's identity through exploiting personal information in varying identity dimensions.

Building on RAT with the novel recognitive proximity approach, we examine the occurrence mechanisms of different types of identity theft. Existing identity theft research has identified five types of identity theft, including credit card fraud, debit card fraud, other existing account fraud, new account fraud and other identity fraud (e.g., tax identity theft, government identity theft) (Anderson et al. 2008; Newman and McNally 2005a). However,

researchers have disagreed on whether all of them are the same crime and should be classified as identity theft. Some studies conceptualize a victimization event as identity theft if any type of these frauds happens to a victim (Goel 2019; Lai et al. 2012; Williams 2016). Some studies focus on credit card fraud and debit card fraud and combine them as a measure of identity theft (Reyns 2013; Reyns and Henson 2016). However, some studies exclude these existing account frauds, and argue that only new account fraud should be classified as identity theft because it can cause more damage and it is harder for the victim to recover from it (Eisenstein 2008; Roberds and Schreft 2009). Equating identity theft to misuse of *any* PII to commit frauds (ITADA 1998) makes it difficult to systematically differentiate these crimes, because each crime involves misuse of PII to some extent. We argue that identity theft involves the *impersonation* of a victim's identity by exploiting personal information in the four identity dimensions, not just misuse of any PII to commit frauds. The closeness of impersonation is determined by the level of recognitive proximity between a criminal and a victim's identity. We distinguish the five types of frauds by mapping the required level of recognitive proximity to each fraud and determine how closely a criminal has to impersonate a victim's identity for each fraud.

Regarding the mechanisms of identity theft occurrence, we consider three different channels that potential offenders can gain unauthorized access to someone else's personal information. Personal information is stored and transmitted in both digital and physical media (Lai et al. 2012). Therefore, we distinguish 1) individual digital channel where individual supplies personal information in online activities such as online shopping; 2) individual physical channel where individual's personal information can be stolen through physical theft. Moreover, personal information are widely distributed among individuals and organizations (Anderson et al. 2008). As a result, the third channel where potential criminals may gain unauthorized access to personal information is through organizational

data breach. By distinguishing the three channels of personal information leakage, we explain the different levels of recognitive proximity created by online proximity and physical proximity to PII through individual and organizational channels. Building on our extended RAT, we develop hypotheses to explain how each channel impacts the risk of each type of identity theft victimization.

We test our hypotheses using the National Crime Victimization Survey (NCVS) and the Identity Theft Supplement (ITS) data. The data is collected from a representative sample of U.S. citizens across three years in 2012, 2014 and 2016. Based on the original sample, we construct a panel data sample with 23,376 observations for data analysis. Our results show that individual online shopping activity increases the probability of credit card, debit card, other existing account fraud and new account fraud. However, individual online shopping activity has no effect on identity theft. It supports our hypothesis that online proximity created by individual online shopping activity increases the probability of frauds that require low recognitive proximity. It is interesting that our results also show that individual physical theft victimization increases the probability of credit card, debit card, other existing account fraud and new account fraud. Opposite to the existing knowledge that online proximity created by individual online activities is the main source of identity theft victimization, this finding show that physical proximity created by individual physical channel is of the same importance. This result provides insight that both online and physical proximity to individual's personal information can lead to to recognitive proximity. In addition, our results on individual's experience in organizational data breach show that organizational data breach increases the probability of all types of identity theft. The results implies that policy makers may increase organization's responsibility for identity theft. This study advances security of personal information and identity theft research by understanding how different channels of personal information

leakage lead to varying levels of recognitive proximity and the corresponding impact on different types of identity theft.

Chapter 2: Theoretical Background

Information systems and criminology research has examined the antecedents of identity theft in light of the routine activity theory (RAT). RAT assumes that criminals are rational actors who make choices of criminal activities based on the embedded social situation (Clarke and Felson 1993). The central statement of RAT hypothesizes that criminal acts become viable options in social situations where likely offenders, suitable targets and the absence of capable guardians converge (Cohen and Felson 1979). This statement characterizes the three core concepts of RAT, including proximity to potential offenders, target suitability and absence of capable guardianship (Bennett 1991). In this study, we focus on developing the concept of proximity in the context of identity theft.

The original RAT was developed to explain the occurrence of street crimes (e.g., larceny theft, burglary). Street crimes occur in the physical space, proximity is defined as the nearness of physical locations between potential offenders and suitable targets (Cohen and Felson 1979). Being physically proximate provides exposure of the target to the potential offenders. According to RAT, the routine organization of individual's everyday activities creates conditions in which the physical proximity to potential offenders will increase. Previous research finds that certain individual routine activities (e.g., go out at night for leisure activities) can increase the physical proximity between potential offenders and suitable targets, thus increasing the opportunity of physical crimes such as burglary and personal theft (Mustaine and Tewksbury 1998).

Since individuals have shift many routine activities to the digital space (e.g., online shopping, online social networking) and there is a rapid growth of cybercrimes, RAT has

recently been extended to explain the occurrence of cybercrimes. Unlike street crimes, cybercrimes can be committed from a long distance which does not require direct physical contact between potential offenders and the targets. Therefore, physical proximity becomes not necessary for cybercrimes to occur. To extend RAT to explain the occurrence of cybercrimes, researchers have proposed the concept of online proximity (Eck and Clarke 2003b) as a substitute for physical proximity. Online proximity is defined as the co-existence of potential offenders and suitable targets within a shared network for an overlapped period of time (Reyns and Henson 2016). Being online proximate provides online exposure of the individual or the individual's personal information to the potential offenders. Researchers have applied RAT with this online proximity approach to examine the link between individual online activities and cybercrime victimization. Some studies find that individual online shopping behavior is positively associated with online consumer fraud targeting and victimization (e.g., retail sales fraud, investment/insurance fraud) (Holtfreter et al. 2008; Pratt et al. 2010). Other studies finds that individual online social networking activity (Henson et al. 2011) and instant messaging usage (Reyns et al. 2011) predict interpersonal online victimization such as harassment and cyberstalking.

2.1 Proximity and Occurrence of Identity Theft

Previous identity theft research has also applied RAT with the online proximity approach to examine individual online shopping behavior and identity theft victimization. This stream of research implicitly assume that mechanism of identity theft occurrence is similar to other cybercrimes: Individual online shopping activities increase the online proximity between potential offenders and suitable targets. Previous research argue that the suitable target in identity theft is PII and the Internet increases ease of availability of personal information online (Goel 2019). However, we argue that identity theft is a

different crime than other cybercrimes. Instead of merely PII, the target is someone else's identity. Therefore, the assumption for RAT with the online proximity approach to be applicable in explaining occurrence of identity theft can be invalid. Our argument is supported by two observations of theoretical gaps in the existing identity theft literature.

First, the findings on the link between individual routine online shopping activities and identity theft victimization have been mixed. Some studies find that individual's online shopping activities increase the risk of existing credit card and debit card frauds (Reyns 2013; Reyns and Henson 2016). On the other hand, one study finds that individual's online shopping activity has no effect on more severe types of identity theft (e.g., filing fraudulent tax return, applying for government benefits or a job), although it increases the risk of existing credit and debit card frauds (Burnes et al. 2020). As individual online shopping activities increase online proximity between potential offenders and potential victim's personal information for all types of identity, these findings suggest that RAT with the online proximity approach can only explain credit and debit card frauds, but not some severe types of identity theft. These findings also suggest that while credit and debit card frauds may be similar to other cybercrimes such as consumer online frauds, the severe types of identity theft can be different crimes with different occurrence mechanisms. Some researchers have also noted that an individual's identity is stolen in severe types of identity theft such as opening new accounts and applying for a job, but not in credit and debit card frauds (Eisenstein 2008).

Second, while previous identity theft research has equated leakage of PII through individual online shopping activities with identity theft, online proximity may be neither a necessary nor sufficient condition for identity theft victimization. Although identity theft has become the fastest growing crime in recent years, it has existed before the Internet Era (Newman and McNally 2005b). Apart from obtaining access to sensitive personal

information from the Internet, criminals can also access it through low-tech means such as burglary, mail theft, pickpocketing, and stealing wallets or purses (Newman 2004a). Therefore, identity theft is not a cybercrime although its occurrence can be assisted by computer networks. Having online proximity is not a necessary condition for identity theft. However, to our knowledge, no study has examined the mechanism of identity theft other than the online proximity approach.

In addition, obtaining sensitive personal information from the online channel is not sufficient for identity theft. However, sufficiency is often implicitly assumed in existing identity theft research as reflected in the wide adoption of the statutory definition of identity theft in identity theft research, which defines identity theft as the misuse of *any* PII to commit illegal activities for fraudulent purposes (ITADA 1998). In contrast, some scholars argue that PII lost does not necessarily mean PII abused subsequently for identity theft crimes (Mann 2015). Some scholars also state that "the probability distributions of actual abuse of stolen information conditional on a breach, are, if not unknown, extremely uncertain" (Acquisti et al. 2016).

An example that illustrates that online proximity is neither a necessary nor sufficient condition for identity theft victimization is the case of Frank W. Abagnale Jr., who was a 'successful' con man later turned into a security consultant (Abagnale 2013). Abagnale conducted identity theft in the 1970s before the Internet existed. He obtained someone else's information through the physical channel and observation of someone else in-person. He not only forged personal checks using stolen check information, but also impersonate a pilot, a doctor and an attorney by exploiting additional information and knowledge to behave and speak like these roles. This case vividly shows that identity thieves not only forge checks, but also become imposters by "stealing other people's lives."

We build on RAT to further understand the occurrence mechanisms of identity theft. We also depart from previous studies in the approach of conceptualizing proximity in the context of identity theft. We define identity theft as the criminal acts of *impersonating* someone else by misuse of his/her personal information – not just obtaining and misuse of any PII – to commit illegal activities for fraudulent purposes. We propose a novel concept of recognitive proximity between a potential offender and an individual's identity that measures how closely the offender impersonate the victim, which in turn determines the probability of identity theft victimization. In the next section, we first briefly review different theoretical forms of proximity and discuss the uniqueness of proximity to identity. We also clarify the definitions and dimensions of identity. We then use this theoretical framework as the foundation to develop our novel concept of recognitive proximity and clearly explain the meaning of closeness of impersonation.

2.2 RECOGNITIVE PROXIMITY IN IDENTITY THEFT

While the concepts of physical proximity and online proximity have been studied in RAT, academic researchers have also studied other forms of proximity. For example, organizational researchers have summarized three main forms of proximity, including 1) physical proximity, 2) functional proximity (Moodysson and Jonsson 2007) and 3) psychological proximity (Ghorbani et al. 2013; Monge et al. 1985). The physical proximity conceptualize proximity as a simple linear distance between people where the physical distance acts as a natural barrier of interaction between people. The conventional physical proximity concept in RAT is consistent with this form of proximity. In contrast, researchers who hold the functional proximity view argue that proximity between people is affected by both the physical space and the functional facilities in it. For example, number of transportation facilities (airlines, trains) can better predict travel route than absolute

physical distance between two locations. The online proximity view in extended RAT is consistent with this form of proximity where the shared network serves as the functional facility that determines the proximity of information to potential offenders online. Rather than focusing on linear distance between people or physical distance altered by functional facilities, the psychological proximity examines the psychologically perceived distance between people. For example, friends may perceive themselves to be close even when they are located in distant physical locations. Recent research has also developed new forms of proximity, including social proximity and cognitive proximity (Molina-Morales et al. 2014), and shows that they are better predictors for innovation collaboration between organizations than spatial proximity (von Proff 2016). Table 1 summarizes the above forms of proximity based on brief review of literature related to proximity.

Table 1. Definition of varying forms of proximity in a selection of studies.			
Form of proximity	Definition	Application research context	Studies using this definition
In the RAT literature			
Physical proximity	Physical distance between two actors.	The two actors are the offender and the target individual; physical proximity explains the probability of street crime victimization.	Cohen and Felson 1979; Mustaine and Tewksbury 1998.
Online proximity	Co-existence of two actors in a shared network.	The two actors are the offender and the target individual or target PII; online proximity explains probability of cybercrimes such as online frauds and cyberstalking.	Reyns and Henson 2016; Henson et al. 2011; Pratt et al. 2010; Holtfreter et al. 2008
In other social science literature			
Functional proximity	Physical distance affected by mobility or accessibility (e.g., transportation facility) between two actors.	The two actors are two organizations seeking collaboration. Functional proximity facilitates knowledge collaboration between two organizations.	Moodysson and Jonsson 2007
Psychological proximity	Perceived psychological closeness (e.g., emotions, feelings) between two persons.	The two actors are an individual offender and a victim. Psychological proximity influences the offender's level of perceived guilt, shame, and compensation to the victim.	Ghorbani et al. 2013
Social proximity	Closeness of relationship based on knowledge about each other and mutual friendship between two actors which results in trust.	The two actors are two organizations. Social proximity predicts the probability of innovation collaboration of small and medium-size organizations.	von Proff 2016
Cognitive proximity	Cognitive homogeneity between two actors, including shared language, common norms and values.	The two actors are two organizations. Cognitive proximity facilitates innovation of an organization through knowledge acquisition from cognitively proximate organizations.	Molina-Morales et al. 2014

Table 1: Definition of varying forms of proximity in a selection of studies.

According to RAT, the concept of proximity aims to explain how close the suitable target is from the potential offenders. Given target suitability and capable guardianship, the closer the suitable target is from the potential offenders, the more likely the crime would occur. For identity theft, the target is someone else's identity. Therefore, our goal is to understand the meaning of being close to someone else's identity. Rooted in social psychology theory, identity is the multifaceted and complex answer to the philosophical question "Who am I?" (Erikson 1968; Ma and Agarwal 2007). Identity involves three actions including knowing, claiming and recognizing (Chryssochoou 2003). As for

knowing about oneself, identity is defined as " the individual's self-appraisal of a variety of attributes along the dimensions of physical and cognitive abilities, personal traits and motives, and the multiplicity of social roles including worker, family member, and community citizen" (Whitbourne and Connolly 1998). Based on this self-knowledge, individuals can make claims to others about themselves. Others also use this information for recognition of an individual, i.e., answering the question "Who are you?" (Chryssochoou 2003).

For an identity thief, the first step towards stealing someone else's identity is to know the target identity by collecting personal information. However, the goal of an identity thieve is not to cognitively appraise himself/herself in the same way as the target identity and become cognitively proximate. The identity thieve still knows that he/she has own identity that is different from the victim's identity. Instead, the identity thief's goal is to impersonate the victim by assuming his/her identity and be recognized as the victim in interactions with others. Therefore, we propose the concept of recognitive proximity to measure how closely an identity thief impersonates the target identity so that he/she can be recognized as the victim. While it is intuitive to roughly understand what impersonation means by imagining an imposter who pretends to be someone else, it becomes more intricate when using it to define recognitive proximity. To systematically develop the concept of recognitive proximity, we build on the four-dimensional framework of identity and use it to explain the details of recognitive proximity.

2.3 Identity and Its Four Dimensions

Researchers have demonstrated that individual's identity falls into four dimensions. These include personal identity, role identity, social identity and symbolic representation (Camp 2004; Craig et al. 2019). Personal identity is the unique set of attributes on

characteristics and traits of an individual (Ashforth 2001). It includes biometric characteristics about an individual's body, patterns of an individual's behavior, history of activities, and psychological characteristics about an individual's personality traits and beliefs (Eccles 2009; Poster 2007). These beliefs and characteristics comprise the self as a unique and idiosyncratic individual that is distinct from others (Craig et al. 2019). In the e-commerce context, merchants can use individuals' behavioral data such as purchasing and browsing history to infer individuals' tastes and preferences (Acquisti 2008). This is reasonable because individuals behave in predictable ways that can preserve personal identity continuity, which give them a harmonious experience aligned with an "authentic self" (Craig et al. 2019; Iyer and Jetten 2011).

Role identity refers to the dimension of identity related to skills and competencies that comprise a role the individual plays in social contexts (Eccles 2009; Stryker and Burke 2000). The role identity is accompanied by a conceptualization of what it means to be "X" (e.g., a doctor) (Petriglieri 2011). It informs individuals' own perceptions and others' expectations for success and importance individuals attach to tasks. For example, an individual who works as a doctor has a doctor role identity. A doctor may associate the meanings of "skillful", "expertise", and "caring" with this professional role. These meanings help others establish the beliefs about how well the individual perform the tasks associated with the role of doctor (Burke 1991). They also become part of how an individual views the self (Anteby 2008).

Social identity refers to the collective beliefs about the self as the member of a group (Craig et al. 2019). It is the dimension of identity that is related to one's ties to highly valued social groups and relationships — such as one's gender, race, religion, social class, culture, and family (Eccles 2009). It helps individuals define the "WE" self, both for themselves and for the people who they interact with. An individual's attributes that

confirm beliefs and goals of a group (being ingroup) raise the individual's value for that group (Craig et al. 2019). In contrast, attributes that disagree with beliefs and goals of a group (being outgroup) reduce this form of value. For example, belonging to the group of children, or unemployed may be valued by people whose goal is to protect and help the under privileged people.

The personal identity, role identity, and social identity communicate substantive information about an individual in three different dimensions. Personal identity relates to basic facts about an individual, role identity relates to meanings related to specific roles in social contexts, and social identity relates to values based on social group membership. Instead of being substantive, the fourth dimension of identity is symbolic representation of a human being (Duck and McMahan 2021). The most familiar symbolic representations are first name and last name of an individual. In some situations, it is suitable for an individual to remain anonymous when interacting with others, e.g., buying something at a kiosk using cash. But in many situations, individuals use symbolic representations with organizations so that they can be memorized and recognized in future interactions. Some common public identifiers are used as symbolic representation of an individual across different contexts, such as name, government ID and social security number. Some identifiers are only used for representing an individual within its own context, such as credit card number issued by a credit card company, which is also referred to as pseudonym (Camp 2004).

2.4 Scale of Recognitive Proximity

Previous research in Information Systems and Information Security suggests that recognition process is a combination of two steps including identification and authentication (Glynos et al. 2005; Zviran and Erlich 2006). We distinguish machine

recognition done by computer systems from human recognition. We also differentiate the level of recognitive proximity required in varying recognition processes. Depending on the context and the interacting party, an individual can have many different identities, which are also referred to as partial identities (Claub and Kohntopp 2001). Figure 1. illustrates the variation of the scope of personal information and identity dimensions involved in recognizing varying partial identities.

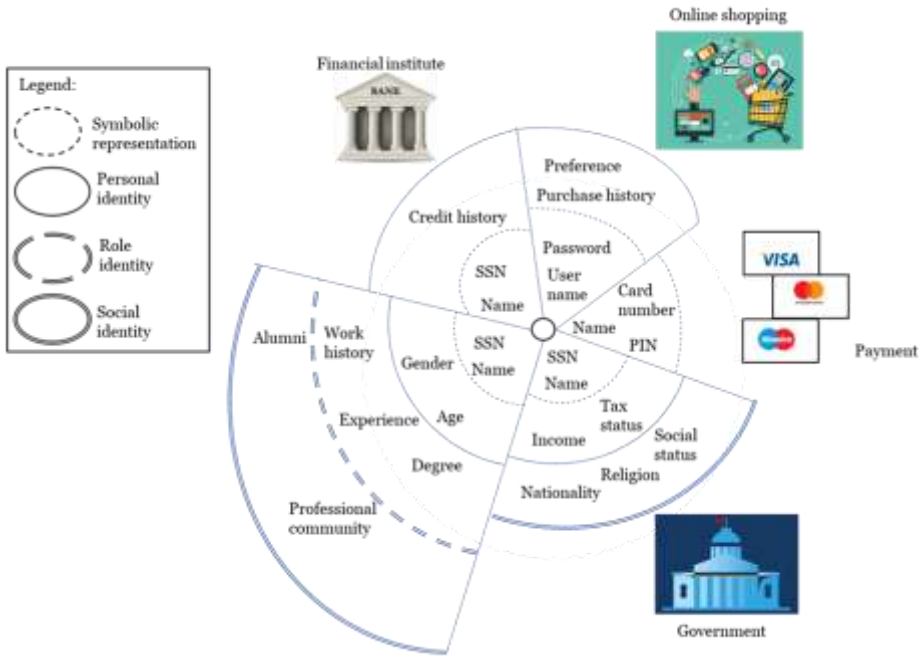

Figure 1: Scopes of identity dimensions and PII for varying partial identities.

The first step in a recognition process is identification. Identification is defined as the association of PII with a unique person (Camp 2004). It answers the question "Who are you?" – Individuals supply a set of symbolic representation information as PII, such as name, social security number, and username to computer systems or other people for claiming who they are and differentiating from others. The second step in a recognition

process is authentication. Authentication is defined as proof of an information (Camp 2004). Since anyone can supply fake PII and claim as someone else, authentication is aimed to verify this claim and increase the confidence of recognition.

As shown in Figure 1, some partial identities of an individual can be recognized by computer access management systems, where identification and authentication are done by computing machines. The goal of the access management systems is to recognize authentic users and provide them with access to resources. Traditionally, authentication methods used in machine recognition processes are divided into three types (Steinbart et al. 2016; Zviran and Erlich 2006): a) what an individual knows (e.g., password, PIN, question-and-answer like 'what is your first pet's name?', dynamic passcode), also referred to as knowledge-based authentication. The information supplied by an individual can be either PII (e.g., password) or non-PII (e.g., pet's name); b) what an individual has (e.g., smart card tokens, certificate, signature), also referred to as possession-based authentication. It requires private objects that an individual possesses; c) what an individual is (e.g., fingerprint, facial scan, retina scan), also referred to as biometric-based authentication. Recently two authentication methods are emerging (Young 2020): d) what an individual does (e.g., keyboard dynamics as how one types); and e) where an individual is (e.g., geolocation).

For an identity thief to impersonate someone else's identity and be recognized as the victim, at least some information in one identity dimension has to be exploited. It is clear that if none of personal information in any of the four identity dimensions is exploited for impersonation, recognitive proximity is zero between a potential offender and a target identity. For machine recognition processes that require no authentication, or single factor knowledge-based authentication, impersonation only requires PII in the symbolic representation dimension. This is the type of impersonation that requires the least scope of

identity dimension, and it has low recognitive proximity. For machine recognition processes that require two-factor authentication involving biological and behavioral characteristics (Glynos et al. 2005), impersonation requires both PII in the symbolic representation dimension and information in the personal identity dimension, and it has medium recognitive proximity.

Human recognition process is even more sophisticated than machine recognition process. A human being recognizes another person by combing perceptions of visual information, aural information, as well as scent, touch and behavior (Glynos et al. 2005). As in Frank Abagnale's case, he impersonated a pilot, a doctor, and an attorney by dressing, talking and behaving like these roles in addition to provide PII on paper (Abagnale 2013). In addition to provide access to goods or services, human recognition of identity is also associated with generating trust about the competency and social status of the individual (Voci 2006). This type of impersonation requires all four identity dimensions, and it has high recognitive proximity. At the same time, human recognition also works together with machine recognition in situations where medium to high recognitive proximity needs to be verified. For example, a cashier can verify the face on a photo ID of a customer paying with a credit card, where both symbolic representation and personal identity information are authenticated. A hiring manager can utilize resume filtering software and conduct face-to-face interview with job applicants to verify all identity dimensions of the applicant.

Based on the above reasoning, we define recognitive proximity as the closeness of impersonation by exploiting varying scope of personal information in identity dimensions of symbolic representation, personal identity, role identity, and social identity between an imposter and a victim resulting in recognition as the same partial identity, which provides access to resources, or trust from others. This definition suggests that recognitive proximity reaches from not exploiting personal information in any identity dimensions (zero

recognitive proximity) to impersonation using information in all identity dimensions (high recognitive proximity). Figure 2 summarizes the scale of recognitive proximity.

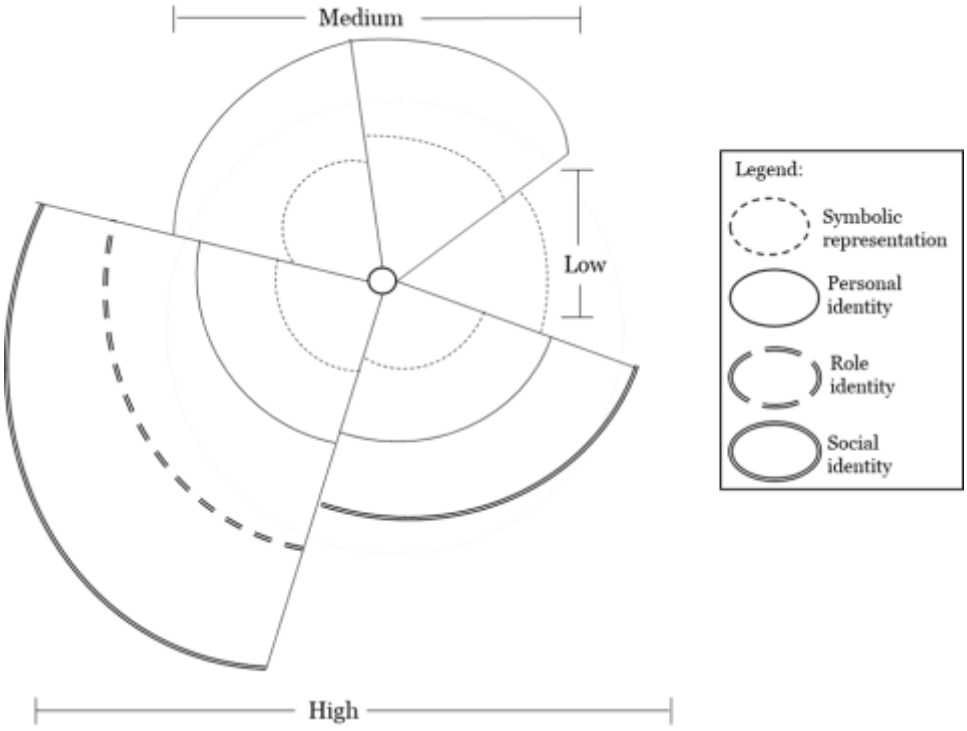

Figure 2: Scale of recognitive proximity.

As Figure 2 shows, the scale of recognitive proximity reflects different levels of closeness an identity thief impersonates the identity of a potential victim. However, the widely adopted statutory definition of identity theft in previous research does not fully distinguish levels of recognitive proximity. It views identity theft as the misuse of *any* PII to commit illegal activities for fraudulent purposes (ITADA 1998). PII is in the symbolic representation identity dimension, it is involved in all levels of impersonation from low to high recognitive proximity. In other words, the existing statutory definition of identity theft lump together scales of recognitive proximity that are greater than zero. In comparison, our

approach advances the understanding to identity theft by distinguishing the levels of recognitive proximity. In the next section, we further distinguish different types of fraud that are categorized as identity theft in previous research and show that they are different crimes.

2.5 CONCEPTUALIZATION OF IDENTITY THEFT

Table 2 summarizes the definitions and measurements of identity theft in a sample of existing identity theft research. As shown in Table 2, existing identity theft research commonly measure identity theft with five types of fraud, including existing credit card fraud, existing debit card fraud, other existing account fraud, new account fraud, and other identity theft. These types of fraud all fall under the definition of identity theft because they involve misuse of *any* PII for fraudulent purposes.

Study	IDT definition	IDT measurement (if applicable)	Scope of personal information
Burnes et al. 2020	The intentional, unauthorized use of a person's identifying information (PII) for unlawful purposes.	Three distinct types: 1) Existing credit and debit card; 2) New account; 3) Other identity theft.	PII
Bose and Leung 2019	Misuse of another individual's information to commit fraud.	*NA*	PII and non-PII
Reyns and Henson 2016	The collection and use of an individual's personal information,without his or her consent, for criminal purposes	Credit and debit card fraud	PII and non-PII
Lai et al. 2012	A crime in which an impostor obtains key pieces of personal identifying information (PII) such as Social Security numbers and driver's license numbers and uses them for their own personal gain.	Mixture of: 1) Existing credit and debit card; 2) New credit and bank account; 3) Apply for government benefit.	PII
Copes et al. 2010	The misuse of another individual's personal information to commit fraud.	Three distinct types: 1) Existing credit card; 2) Existing debit card; 3) New account.	PII and non-PII
Anderson et al. 2008	Acquiring enough data about another person to gain access to a specific account or credit history for unlawful gain.	Any type of: 1) Existing credit card; 2) Other existing account; 3) New account.	PII and non-PII
Eisenstein 2008	The misuse of another individual's personal information to commit fraud.	New account fraud. (A thief has stolen one's identity)	PII and non-PII
Kahn and Roberds 2008	The malicious use of personal identifying data	1) Existing account fraud; 2) New account fraud.	PII

Table 2: A review of identity theft definitions in a sample of studies.

However, some researchers have argued that credit card fraud should not be classified as identity theft (Copes et al. 2010), and new account fraud and other identity theft are crimes where real identities are 'stolen' (Eisenstein 2008) based on empirical differences in victim profiles. We argue that these are different crimes because they require different levels of recognitive proximity between the potential offender and target identity. Some frauds are not identity theft because identity theft involves the criminal acts of *impersonating* someone else by misuse of his/her personal information – not just obtaining and misuse of any PII – to commit illegal activities for fraudulent purposes. We build on the four-dimension identity framework and use recognitive proximity to conceptually clarify the difference between these crimes. Table 3 presents the identity dimensions and level of recognitive proximity involved in both machine and human recognition processes for these crimes. We explain the details in the following.

	Recognition process	Scope of identity dimensions	Level of recognitive proximity	Result of impersonation
Credit card fraud	Online: machine recognition; Offline: human recognition	Symbolic representation	Low	Access to money and goods
Debit card fraud	Online: machine recognition; Offline: human recognition	Symbolic representation	Low	Access to money and goods
Other existing account fraud	Online: machine recognition; Offline: human recognition	Symbolic representation; Personal identity	Low	Access to money, goods and services
Opening new account	Human recognition	Symbolic representation; Personal identity	Medium	Access to credit and money; Trust from creditors
Identity theft	Human recognition	Symbolic representation; Personal identity; Role identity; Social identity	High	Access to multiple resources; Trust from other people; Perceived social status from other people

Table 3: Categorization of scope of identity dimensions and level of recognitive proximity in identity theft and frauds.

2.5.1 Credit and Debit Card Fraud

Credit and debit card fraud refers to events where criminals impersonate as authorized users to credit and debit card accounts to obtain money or payment services (Cheney 2005). Machine recognition for authorized users to credit and debit card accounts

are widely adopted because of its convenience for online payment and e-commerce. For machine recognition of credit and debit card users in the online environment, supplying credit and debit card information for identification is often sufficient, which includes the card account number, expiration date and the cardholder's name (PCIDSS 2018). The authentication of credit and debit card user in machine recognition requires private PII such as a security code (e.g., CVV2) of the credit card and a PIN of the debit card. Therefore, machine recognitive of credit and debit card users in the online environment only requires PII in the symbolic representation dimension of identity. There is low recognitive proximity between the criminal and the victim's identity in such online credit and debit card fraud.

In the offline environment, human recognition of credit and debit card users are sometimes required. In addition to the symbolic representation identity dimension, biological characteristics in the personal identity dimension is required in human recognition. For example, merchants are required to authenticate the photo ID of a customer using credit or debit card for payment during in-person transactions. Thus, it seems to require medium recognitive proximity between the criminal and the victim's identity in the offline environment. However, one study shows that the percentage of cashiers who check photo ID of a customer is as low as 5% in restaurants and grocery stores, and 0% customer IDs are checked in gas stations (Downing et al. 2016). In addition, purchasing goods and services online is a more convenient substitute to offline transactions. Therefore, we conclude that low recognitive proximity is required for credit and debit card frauds. Since the criminals do not need to impersonate the characteristics in the personal, role and social dimensions of target identity in these frauds, they should be categorized as frauds instead of identity theft.

2.5.2 Other Existing Account Fraud

Although credit and debit card frauds are the most frequently reported crimes regarded as identity theft, researchers have also examined other types of existing account fraud including telephone service accounts, utility accounts, membership accounts and online service accounts (Anderson 2006; Anderson et al. 2008). This type of fraud involves events where criminals impersonate as authorized users to access these existing accounts. Since many merchants and organizations provide services through online websites, machine recognitions for authorized users are often used in the online environment. Individuals use PII such as account number, username, and passwords to access these accounts online. These pieces of PII fall into the symbolic representation identity dimension, while information in other identity dimensions is not required for machine recognition in the online environment for other existing account fraud. Thus, there is low recognitive proximity between the criminal and the target identity.

Obtaining services from other existing account in the offline environment may require additional information in the personal identity dimension. For example, to change a utility service to a new address, proof of both old and new address is required by the utility provider. A utility account user also needs to provide billing history as proof of activities. Similarly, a merchant member (e.g., Sam's Club, Costco membership) needs to provide photo ID to a cashier in the offline environment to obtain the membership service. The cashier may also ask about previous purchase history to verify whether the individual responds appropriately. Therefore, human recognition for other existing accounts in the offline environment requires information in both symbolic representation and personal identity dimensions, resulting in medium level of recognitive proximity between the criminal and the target identity. However, utility providers have also began offering online services (Hamad et al. 2017), and merchants also provide online ordering service for online

purchasing. Therefore, criminals can use the online channel with machine recognition to substitute the offline channel with human recognition. We conclude that low recognitive proximity is required for other existing account fraud. Similar to credit and debit card fraud, this type of crime should be categorized as non-identity theft because the criminals do not need to impersonate the victim's personal, role, or social identities.

2.5.3 New Account Fraud

New account fraud refers to the events where criminals impersonate the victim's identity to open new financial accounts and create credit relationships (Eisenstein 2008). Different from credit and debit card fraud, or other existing account fraud, where criminals is recognized as authorized users and gain access to money or services from existing accounts, creditors recognized criminals as trustworthy individuals whom they can lend money to. Therefore, recognition process requires credit history information in the personal identity dimension, in addition to PII in the symbolic representation dimension of identity, to determine the credit worthiness of an individual. In the U.S., organizations that provide credit are required to know customers' personal identities by regulations of Customer Identification Program (FederalRegister 2003). Authentication involves proof of credit history and income history of the individuals, including government issued ID, social security numbers, and W-2 forms. While W-2 forms serve as direct proof of income history, social security number and government issued ID can be used to link to individual's credit history in credit bureau's (i.e., Experian, Equifax, and TransUnion) database (Eisenstein 2008).

The recognition process for opening new accounts is conducted by human beings. Even in online application of opening a new account, employees from financial organizations review the above information supplied by an individual and then decide

whether to approve the application. In the offline environment, an individual may need to communicate with a clerk, provide live signature, in addition to providing required personal information. The clerk may check whether there is any red flag by observing the individual's behavior, listening to the individual together with examining the supplied information. Therefore, the criminal needs to impersonate the target identity in the personal identity dimension well, in addition to providing PII in the symbolic representation dimension, to be recognized as a credit worthy individual. We conclude that new account fraud requires at least medium level of recognitive proximity. Consistent with previous research, we categorize new account fraud as one type of identity theft (Eisenstein 2008).

2.5.4 Identity Theft

Identity theft refers to the events where criminals impersonate as the victim's identity to interact with non-financial organizations for fraudulent purposes (Anderson 2006). Examples include frauds where criminals impersonate as the victim's identity to apply for a job, file tax returns, obtain unemployment benefit from government, and record as the victim when stopped with a crime by law enforcement. The recognition of these identities involves human beings, such as job interviewer, tax professionals, government employees and law enforcement agents.

Moreover, the criminals gain trust about the identity's competence or social status, in addition to access to money or services, in these frauds. For example, in applying for a job by impersonating the role identity of a victim, a criminal obtains the employer's trust on his/her skills related to the job. In applying for government unemployment benefit, a criminal obtains the government trust on his/her social status of being unemployed and the eligibility for benefits. As in Frank Abagnale's case, his impersonation of professionals as a doctor, a pilot and an attorney makes other people more likely to trust him because of the

social status represented by these professional roles (Abagnale 2013). The criminal needs to supply information in all identity dimensions to gain trust on the social role or social status of the identity. In addition, the criminal may need ability as a con man to behave like the impersonated role in the human recognition processes. Therefore, we conclude that other identity theft requires high level of recognitive proximity, in which the real identities of the victims are 'stolen'. We refer to these types of fraud as 'identity theft' in the remaining text of this paper.

Chapter 3: Hypotheses Development

3.1 Individual Online Shopping Activity and Identity Theft Victimization

Individual online shopping activity is an important online routine activity that creates proximity between potential offenders and suitable targets through shared network. Previous research has built on existing RAT and focused on the concept of online proximity to explain the link between individual online shopping activity and identity theft victimization (Reyns and Henson 2016). We argue that the risk of identity theft victimization is determined by recognitive proximity rather than online proximity. Although online proximity is related to recognitive proximity, previous research has not examined the level of recognitive proximity that is associated with individual online shopping activity.

To determine the level of recognitive proximity created by individual online shopping activity, we analyze the scope of personal information and the corresponding identity dimensions associated with it. During online shopping activities, individuals share information with merchants for two main purposes, including 1) supplying payment information to complete the transactions and 2) informing the sellers about their purchasing preferences (Acquisti 2008). In the U.S., the predominant payment methods for individual

online shopping transactions are card payment and digital payment (e.g., PayPal). A recent report shows that credit card, debit card and PayPal account for 44%, 32%, and 14% of the payment methods for individual online shopping transactions (Chevalier 2021). In addition, one report shows that card payment and digital payment account for 75% of all e-commerce transactions in the U.S. (J.P.Morgan 2021). Recognitions of authorized users to credit and debit and PayPal are conducted by computer systems online. Individuals provide symbolic representation information, including names, credit and debit card numbers, card expiration dates, or alternative digital payment account number (e.g., PayPal), for identification to these payment methods (PCIDSS 2018). Authentication in the machine recognitive process also requires just symbolic representation information, including debit card PINs or digital payment account passwords.

In addition to payment information that are required to complete online transactions, merchants may also track individual's online browsing and purchasing behavior to construct consumer profiles. Merchants have incentives to know more about the consumers' preference by tracking their behavior because they want to implement personalized pricing and service strategies to maximize profit (Acquisti 2004; Acquisti 2008). For example, e-commerce platforms such as Amazon and Ebay allow users to create user profiles that keep a record of users' purchase history. Online shopping websites can also use cookies to track users' browsing behavior (Degeling et al. 2019).

The scope of personal information involved in individual online shopping behavior mainly consist in the symbolic representation and personal identity dimension. As summarized above, online payment information consists in the symbolic representation dimension. Using information in only the symbolic representation dimension, a potential criminal can impersonate an authorized user to the level of low recognitive proximity, as shown in Figure 1. Low recognitive proximity is sufficient for credit and debit card frauds,

and other existing account fraud to be conducted online through machine recognition. The scope of personal information involved in individual online shopping activity may even lead to these frauds in the offline environment. In addition to the symbolic representation information, the online browsing and purchasing history information consists in the personal identity dimension. Although this information is intended for the merchants to know more about their customers, potential offenders can also exploit this information to impersonate customers. For example, in offline purchasing environment, merchants such as Costco and Sam's club require the customers to verify their identities with cashiers. This recognition process involves human recognition. In addition to presenting the stolen membership account usernames, membership cards, knowing the purchase history enables the criminals to communicate such personal activity history to the cashiers. Such impersonation is more recognizable in the human recognition process.

In summary, the scope of personal information involved in individual online shopping activity is mainly in the symbolic representation identity dimension. It is also associated with information in the personal identity dimension but limited to individual's online browsing and purchasing history. Indeed, the most frequently leaked personal information online is credit and debit card information and online account login credentials (Zaiss et al. 2019). Therefore, we conclude that individual online shopping activity creates low recognitive proximity between potential offenders and target identities. Low recognitive proximity is sufficient for credit and debit card fraud, and other existing account fraud. Thus, individual online shopping activity will increase the possibility of victimization of these frauds. However, at least medium recognitive proximity is required for new account fraud, which involves personal identity dimension related to credit worthiness. For identity theft, recognition even requires high recognitive proximity involving the role identity and social identity dimensions. As the scope of personal

information in individual online shopping only involves the symbolic representation identity dimension, and personal identity dimension related to online browsing and purchasing histories, it is not likely to increase the possibility of new account fraud and identity theft victimization. Therefore, we hypothesize that:

H1a. Individual's online shopping activities increase the probability of credit card fraud victimization.

H1b. Individual's online shopping activities increase the probability of debit card fraud victimization.

H1c. Individual's online shopping activities increase the probability of other existing account fraud victimization.

3.2 Individual Physical Personal Information Leakage and Identity Theft Victimization

Rather than online proximity, the original RAT has focused on physical proximity between potential offenders and suitable targets. Physical proximity measures the physical distance between the potential offenders and the suitable targets. However, existing identity theft research has not yet examined the relationship between physical proximity and identity theft victimization. We argue that physical proximity is related to recognitive proximity because physical documents containing personal information is held by individuals. If potential offenders gain physical proximity to these documents, it is possible to increase the recognitive proximity, thus increasing the risk of identity theft victimization. Since our aim is to explain the occurrence mechanism of identity theft victimization, we do not focus on individual physical routine activities that may be antecedents of physical proximity. Instead, we focus on individual experience of physical theft where physical proximity is already created.

Personal physical theft refers to physical crimes involving personal properties, including larceny theft, home burglary and car burglary. When these crimes happen, criminals may gain physical proximity to documents containing personal information. Since an individual is the subject of his/her identity, the paper documents the individual holds may cover every identity dimension. First, many physical ID cards or paper documents contain symbolic representations of an individual that can be used in many contexts. For example, physical credit and debit cards contain names and account numbers that are symbolic representations used in interactions with payment service providers. Merchant membership cards contain names and account numbers that are symbolic representations used in interactions with merchants (e.g., Costco, Sam's Club). Government issued IDs (e.g., driver license, passport, social security card) contain names and ID numbers that are symbolic representations used in interactions with specific government departments. Second, information about one's personal identity is also contained in many documents. For example, birth certificate, photo IDs contain one's biological information, billing statements contain one's address and transaction histories with organizations, income tax forms (e.g., W-2 forms) contain one's income history and employment history. Third, information about one's role identity can be contained in education degrees, job training certificates that show an individual's competence in professional areas. Finally, information about one's social identity can be contained in passport that shows one's nationality, and documents that can show one's belonging to social groups. For example, an employee card of a doctor shows one's belonging to a hospital with the social status of a doctor.

The physical proximity created by larceny theft may create a low level of recognitive proximity. Purse, wallet, backpack, and smartphones are the most common properties stolen in larceny thefts (Mustaine and Tewksbury 1998). These personal

belongings are commonly used to carry credit and debit card, insurance cards and photo IDs for daily usage in addition to cash (Newman 2004b). Even when the main goal of an ordinary thieve is money, the thieve has an incentive to also steal the credit and debit card and other IDs for potential monetary gain by exploiting the information for frauds. This low level of recognitive proximity created in larceny theft is sufficient for credit and debit card frauds, and other existing account fraud. Indeed, the criminals can use the stolen credit and debit cards to purchase goods in person. Although merchants are required to verify the personal identity of a customer using credit card to pay for in-person transactions, one study shows that the percentage of cashiers who check photo ID of a customer is as low as 5% in restaurants and grocery stores, and 0% customer IDs are checked in gas stations (Downing et al. 2016). Comparing to misuse of credit card, misuse of debit card seems more difficult because it may require other information about symbolic representation (e.g., PIN) for in-person authentication. However, one study finds that one out of every eleven stolen wallet leads to misuse of debit card because people tend to use birthdate as the PIN (Bonneau et al.), while birthday information can be found on government issued IDs. In addition, smartphones can contain login credentials to other existing accounts. We hypothesize that:

H2a. Individual physical theft victimization increases the probability of credit card fraud victimization.

H2b. Individual physical theft victimization increases the probability of debit card fraud victimization.

H2c. Individual physical theft victimization increases the probability of other existing account fraud victimization.

Home burglary and car burglary provides the criminals with physical proximity to paper documents containing information in more identity dimensions than larceny theft.

Burglars may gain access to physical documents that contain information about personal, role and social identity. Examples include W-2 forms that contain income history information of personal identity, education records and employment records that contain information about skills and competence related to role identity. In addition, education records and passports also contain information about alumni status and nationality related to social identity. These documents can be found in homes, cars and mailboxes which are designated 'safe' places (Newman 2004b). These places can be breached by burglars, especially those who have relationship with an individual. Security practitioners report that 43% of burglars know their victim (CalderSecurity a), while physical documents containing personal information are highly ranked in burglars' wish list (CalderSecurity b). When burglars are savvy about identity information and know what documents to look for, they can gain medium to high level of recognitive proximity to the target identity. To impersonate the target identity at a medium to high level of recognitive proximity, the criminal may need to also behave like the target identity in the human recognition process. Although not all burglars are sophisticated about exploiting role and social identity information, the medium to high level of recognitive proximity will still increase the probability of new account fraud and identity theft. We hypothesize that:

H2d. Individual physical theft victimization increases the probability of new account fraud victimization.

H2e. Individual physical theft victimization increases the probability of other identity fraud victimization.

3.3 Organizational Data Breach and Identity Theft Victimization

Personal information is widely distributed in organization databases by organizations, in addition to being used and held by individual themselves in online and

physical media. However, existing concepts of proximity in RAT only focus on the forms of proximity between potential offenders and individuals. For online proximity, it focuses on the co-existence of an individual and potential offenders through shared network (Reyns and Henson 2016). For physical proximity, it focuses on the physical distance between an individual and potential offenders (Mustaine and Tewksbury 1998). We argue that recognitive proximity created by the organizational channel is an important antecedent to the occurrence of identity theft victimization. Through obtaining personal information from organizational data breaches, criminals can exploit that information to impersonate target identities without gaining online or physical proximity through individual channels.

The scope of personal information stored in an organizational database is possible to cover every identity dimension of an individual. For example, the Equifax data breach has leaked sensitive personal information in all identity dimensions, since the credit bureau has collected information to construct a detailed portrait of credit history of an individual. It collects information from credit card companies, banks, employers, and other organizations that interact with individuals (EPIC 2018). Leaked personal information from the Equifax data breach include symbolic representation information (e.g., names, credit card numbers, social security numbers), personal identity (credit history), role and social identity (employment history). Even for ordinary organizations, whose goal is not to collect all details of an individual to build a profile of his/her credit history, they try to collect as much information as possible and use it for data analytics to inform business decisions (Acquisti 2008). For example, Target creates IDs in the database for all customers and associate them with the customers' contact information, purchase histories, and other demographic information (Hill 2012). Using detailed customer information with data analytic algorithms, Target was able to find out that a girl was pregnant before her farther did. Therefore, personal information leaked from ordinary organizational data breaches has

the potential for the criminal to gain personal information in all identity dimensions of a victim.

Since the scope of personal information leaked from an organizational data breach can cover every identity dimension, it can facilitate the full spectrum of recognitive proximity from low level to high level. A recent report shows that the top breached PII data types include individual names, social security numbers, username/email and passwords combinations, driver license and other government IDs, and full credit and debit card information (ITRC 2020). These data types include symbolic representations that can facilitate low recognitive proximity required in credit and debit card frauds, and other existing account fraud. This report also shows that more than 60 types of data attributes are leaked in data breaches, covering dimensions of personal identity, role identity, and social identity. Examples of leaked data attributes include W-2 form contents with income history, medical records, education and employment records that are related to role identity, and social group membership information that are related to social identity. The leakage of information in these identity dimensions increase the possibility that a criminal can exploit it to impersonate and be recognized as a victim in human interactions where medium to high recognitive proximity is required, which can result in new account fraud and identity theft. We hypothesize that:

H3a. Individual's experience of organizational data breaches increase the probability of credit card fraud victimization.

H3b. Individual's experience of organizational data breaches increase the probability of debit card fraud victimization.

H3c. Individual's experience of organizational data breaches increase the probability of other existing account fraud victimization.

H3d. Individual's experience of organizational data breaches increase the probability of new account fraud victimization.

H3e. Individual's experience of organizational data breaches increase the probability of other identity fraud victimization.

Chapter 4: Methodology

4.1 Data Sources

This section describes the data used for testing our hypotheses on identity theft victimization. To measure these victimizations at the individual level in the US, we use data from National Crime Victimization Survey (NCVS) and its Identity Theft Supplement (ITS). The respondents of NCVS-ITS are randomly sampled to represent the population of US citizens. One main difficulty in studying identity theft victimization is that data of actual crime events is hard to obtain. As an alternative to measuring actual events, NCVS records victimization data based on respondents' report. NCVS has been used to study crime and victimization in disciplines of criminology (Xie and Baumer 2019), law and economics (Magnus and Steven 2016; Markowitz 2005). Therefore, NCVS and ITS is an important data source to study identity theft victimization.

NCVS-ITS also provides a rich set of survey items that fit well to our research context. It provides separate items asking respondents on distinct identity theft victimization events. These separate items allow us to test crime specific mechanisms that are not studied in previous identity theft literature. ITS also provides items asking respondents on their online behavior, data breach experience, and protection actions against victimization. These items allow us to capture the digital mechanism of identity theft, and account for the impact of guardianship. In addition, NCVS provides items asking respondents on their physical crime victimization, individual and household

characteristics. These items allow us to capture the physical mechanism of identity theft, and account for the risk profile of persons and households.

There are four waves of ITS data available in year 2008, 2012, 2014, 2016. The ITS data can be combined with the annual NCVS data. However, the survey items used in 2008 ITS are not compatible with those used in subsequent ITS. Therefore, we use combined NCVS-ITS data from 2012, 2014, and 2016. The combined overall NCVS-ITS response rates for 2012, 2014 and 2016 are respectively 68.2%, 66.1% and 60.0%. The survey administrators also tested whether respondents and non-respondents differ in important ways, but it did not find any nonresponse bias (Burnes et al. 2020).

We further merge the NCVS-ITS data with three additional datasets based on geographic region information. The first dataset is the Uniform Crime Reporting (UCR) in the US. UCR provides aggregate state level data on all categories of physical crimes, allowing us to capture the overall crime rate in geographic locations. The second dataset is from the Privacy Rights Clearinghouse (PRC). PRC provides organizational level data on various types of data breach events, allowing us to capture the overall data breach rate in geographic locations. The third dataset is from the Current Population Survey (CPS) from the US census bureau. CPS provides aggregate state level data on computer and internet use in the US, allowing us to capture the overall accessibility to computers and internet in geographic locations. Combining these datasets with NCVS-ITS allows us to capture the socioeconomic characteristics of individuals living in different geographic locations in the US.

4.2 Panel Data Sample Construction

Recent research in identity theft has also used NCVS-ITS data (Burnes et al. 2020). They created a pooled cross-section sample for data analysis using logistic regressions.

Their sample construction approach has two potential weaknesses. First, pooling the data across different years takes the assumption that observations are independent. However, NCVS-ITS is a longitudinal survey and the same individual can respond to ITS for more than one time. As a result, the validity of this assumption does not necessarily hold. Second, logistic regression model assumes that there are no omitted variables. If there exist repeated observations for the same individual, we can use fixed effects regressions that rely on less restrictive assumptions. Fortunately, we find it possible to construct a panel data sample from NCVS-ITS although this has not been done in previous identity theft studies.

Table 4. summarizes our sample construction procedure and changes in sample sizes. To construct our panel data sample, we identify four conditions for tracking a respondent. A respondent is trackable if and only if all four conditions hold and will be included in our sample. First, a respondent needs to respond to ITS interviews more than once to allow for repeated observation. One proportion of respondents in the original data does not satisfy this condition because of the rotating panel design of NCVS-ITS. The rotating panel design means that each selected respondent is surveyed for NCVS every 6 months in 3 years and is then rotated out. Since our ITS data are collected every 2 years (in 2012, 2014, and 2016), a respondent can take ITS interviews for a maximum of 2 times before being rotated out. We remove the respondents who respond to ITS for less than 2 times in step-1 as shown in Table 4. Second, given a respondent is not rolled out, the household of the respondent needs to stay at the same home address for the time span of ITS. The reason is that NCVS-ITS uses home address as the tracking information for subsequent interviews. Respondents within a household become not trackable if they move out of the original home address. We remove these respondents in step-2 as shown in Table 4. Third, a respondent needs to remain in the same household for the time span of ITS. Since the NCVS-ITS tracks home addresses instead of persons, a respondent becomes not

trackable if (s)he leaves the original household. We remove these respondents in step-3 as shown in Table 4. Finally, a respondent needs to take ITS interviews by herself/himself. NCVS-ITS allows for a proxy respondent if the respondent herself/himself is not available at the time of interview. Response provided by a proxy can be unreliable (Jeffrey 1988). We remove these respondents in step-4 as shown in Table 4. The above procedure creates our initial panel data sample with 77,738 observations. It enables us to take advantage of the longitudinal NCVS-ITS data and address the weakness of previous identity theft studies.

We further adjusted the initial panel data sample by dropping observations that have missing values in the study variables as shown in step-5 in Table 4. In addition, we remove observations of respondents that report inconsistent values for time invariant variables across years as shown in step-6 in Table 4. The sample size for our final panel data sample used in data analysis is 12,376.

Data extraction criteria	Remaining sample size
Original 2012, 2014 and 2016 NCVS-ITS data	296,575
Drop observations with non-trackability due to non-availability for interview in two waves of NCVS-ITS (2012 and 2014 combination or 2014 and 2016 combination)	84,316
Drop observations that become non-trackable due to moving out of the original household address tracked by NCVS-ITS	83,544
Drop observations who become non-trackable due to changes in household composition (e.g. someone leaves marriage, etc.)	83,290
Extract observations who completed the NCVS-ITS interviews by himself/herself. Drop observations whose NCVS-ITS survey responses are provided by proxy in order to increase response credibility	77,738
Listwise delete observations that has missing values in study variables	12,376

Note: Individuals greater than 16 years are elegible for NCVS-ITS survey interview

Table 4: Panel data sample construction procedures.

4.3 Assessment of Recall Bias

One concern on respondents' victimization reporting is that respondents' recall from memory can be biased. We assess recall bias in the NCVS-ITS and find that the survey is systematically designed to minimize recall bias. The design of interview strategy of NCVS-ITS is based on theories of the cognitive processes involved in responding to retrospective surveys (Groves 2004; Sudman et al. 1996). According to the model of survey response process, the respondent must comprehend the meaning of survey question before retrieving information from memory (Norbert 1995). To accomplish this, a rich set of specific cues to the attributes of crime events (e.g., criminal acts, locales, and relationship to the offender) are included in interview questions. It enhances comprehension by clearly conveying the type of events and avoids the use of legalistic terms (e.g., "identity theft",

"burglary", "victimization"). After comprehension, relevant memory must be accessible for a respondent to retrieve target information (Gabriel and Dipankar 1983). The design also increases the accessibility of memory for information retrieval by making attributes of crime events salient (Deborah and Bernard 1989).

In addition to interview design, the design of data collection of NCVS-ITS also helps decreasing recall bias. The survey data collection technique is called "bounding" (Lauritsen 2001). This means that in the ideal situation, each respondent is tracked and interviewed multiple times before being rotated out. Bounding can generate cross-reference points for the interviewer to prevent duplicate reporting (Bachman and Taylor 1994). It can also generate internal cognitive reference points for the respondent to prevent telescoping (Addington 2005; Lauritsen 2001). In implementing the survey data collection, the NCVS-ITS uses computer assisted and centralized facilities. It enforces the interviewers to follow the survey structure as designed and improve supervision of the interviewers (Cantor and Lynch 2005). These features help to decrease human errors that are not intended by the survey design during the implementation.

We also empirically assess recall bias in our panel data sample. Only 42 cases (<0.4%) reported that they cannot remember enough details about victimization experience in our sample. This is evidence that the survey design is effective. We also check the reporting of some demographic characteristics that are expected to be time invariant (e.g., race) or have a fixed change (e.g., age) for a given respondent. Only 1.6% of respondents give inconsistent reports and they are excluded from the sample[1]. In addition, the correlation between reporting on IDT victimization outside the past 12 months in 2014 (or

[1] The results are similar when these observations are included in the data analysis.

2016) and reporting on IDT victimization in 2012 (or 2014) is 0.20 (p = 0.00). This evidence indicates that respondents' recalls on IDT victimization are not easily forgotten.

4.4 MEASURES

Here we describe the measures of variables used for our main baseline tests.

4.4.1 Dependent Variables

We measure identity theft victimization by whether a respondent has experienced a specific criminal event. Specific criminal events in our study include the following five types: 1) credit card fraud, 2) debit card fraud, 3) other existing account fraud (e.g., telephone account, utility account), 4) new account fraud, and 5) other identity fraud (e.g., PII was used to apply for a job, government benefit, or file tax returns). These types of fraud are consistent with the frauds identified in previous research (Anderson et al. 2008; Burnes et al. 2020; Newman and McNally 2005a) and legal acts (e.g., The Identity Theft Assumption and Deterrence Act of 1998). They are also frequently used by practitioners in the financial industry and government regulators (Cheney 2005).

However, previous literature has not reached an agreement on the classification of these measures. One group of existing studies has classified all of these criminal events as identity theft (Newman and McNally 2005a) because they involve misuse of PII in common. Empirical research in this group has combined them into one measure of identity theft if any type of these fraud occurs to a victim (Lai et al. 2012). On the other hand, researchers have also classified these measures as distinct variables (Anderson 2006; Burnes et al. 2020; Copes et al. 2010). For example, one study finds that credit card fraud, bank account fraud, and new account fraud significantly differs from each other in terms of victim characteristics (Copes et al. 2010).

We argue that these five types of fraud should be studied separately. Although all five types of fraud involve the misuse of PII, the PII attributes may associate with different identity domains. To examine whether there exist underlying common factors that drive these frauds, we also conduct an exploratory factor analysis (EFA). The correlation matrix of these five frauds is presented in Table 5.

	Credit	Bank	OtherEx	NewAcct	IDT
Credit	1.00				
Bank	0.12	1.00			
OtherEx	0.06	0.09	1.00		
NewAcct	0.07	0.07	0.07	1.00	
OtherIDT	0.03	0.02	0.03	0.10	1.00

Table 5: Correlation of financial frauds and identity theft.

The correlations between any two frauds are low. Among all pairs, the credit card fraud and bank account fraud have the largest correlation of 0.12, but the absolute value of correlation is still small. The results in Table 5 give a preliminary indication that these frauds are not driven by any underlying common factors, suggesting that they should be treated as separate variables.

To examine whether the variables share variance due to common underlying factors, we compute the Kaiser-Meyer-Olkin (KMO) measure of sample adequacy based on the correlation matrix (Bandalos 2018). The KMO ranges from 0 to 1. The logic of KMO is that if the variables are highly correlated due to common variance, the partial correlation between two variables after removing effects from other variables should be

low, resulting in a KMO ratio close to 1.0. The KMO statistics for the overall correlation matrix and each fraud variable are presented in Table 6.

KMO for each item					
	Credit	Bank	OtherEx	NewAcct	IDT
	0.57	0.56	0.58	0.56	0.54
KMO for the overall correlation matrix					
	0.56				

Table 6: Kaiser-Meyer-Olkin (KMO) Measure of Sample Adequacy.

According to (Kaiser 1974), the minimum KMO for indication of underlying common factors is 0.6. Our results in Table 6 show that the KMO statistics for the overall correlation matrix and each variable are all below 0.6. Thus, it provides further indication that the five frauds are separate variables.

A principal component analysis (PCA) in the correlation matrix also supports our view to treat the five frauds as distinct variables. Table 7 summarizes the eigenvalues and percentages of variance accounted for by each component from the PCA. In the situation where the components are independent from each other, the eigenvalues for each component will be equal to 1.

Total variance explained			
Component	Eigenvalue	% of variance	Cumulative % of variance
1	1.26	25.29	25.29
2	1.03	20.61	45.90
3	0.94	18.89	64.79
4	0.89	17.73	82.52
5	0.87	17.48	100.00

Table 7: Percentages of Variance Accounted for in 5-Component PCA.

Results in Table 7 shows that all 5 eigenvalues are close to 1, and the variance accounted for by each component are similar. These results suggest that number of components should not be reduced. In addition, rotated factor loadings based on the results of PCA show that each fraud variable loads onto one distinct component with loadings of 1. These results provide consistent evidence that the five fraud variables are measuring different constructs. Therefore, we treat the victimization of each type of fraud as a dependent variable. Each dependent variable is operationalized as whether a respondent experience victimization of the corresponding fraud in that year (2012, 2014, and 2016). Table 8 presents the details of operationalization and summary statistics for the dependent variables in our regressions.

Variables	Operationalization	Mean	SD
Dependent variables			
Credit card fraud	Indicator of whether a respondent experienced victimization of credit card fraud	0.0417	0.2000
Debit card fraud	Indicator of whether a respondent experienced victimzation of debit card fraud	0.0295	0.1692
Other existing accounts fraud	Indicator of whether a respondent experienced victimization of other existing account fraud (e.g. utility and telephone account)	0.0053	0.0725
New account fraud	Indicator of whether a respondent experienced victimization of misuse of PII to open new accounts	0.0040	0.0630
Other identity fraud	Indicator of whether a respondent experienced victimization of misuse of PII for other fraudulent purposes (e.g. getting government benefits, a job or medical care)	0.0028	0.0528
Independent variables			
Individual online shopping	Indicator of whether a respondent has purchased something online	0.5259	0.4993
Online shopping using credit card only	Indicator of whether a respondent uses only credit card for online shopping	0.3017	0.4590
Online shopping using debit card only	Indicator of whether a respondent uses only debit card for online shopping	0.0983	0.2977
Individual physical theft	Indicator of whether a respondent experienced any of larceny theft, home burglary or car burglary	0.0515	0.2210
Organizational data breach	Indicator of whether a respondent has received data breach notification letters from organizations	0.0912	0.2879
Organizational data breach containing SSN	Indicator of whether a respondent has received data notification letters indicating SSN was included in the breached information	0.0207	0.1425

Table 8: Variable Operationalization and Descriptive Statistics

Now we describe our hypothesized independent variables and other control variables that could impact the probabilities of the five-fraud victimization. We include these variables in the analysis based on the theoretical foundation of routine activity theory. In specific, routine activity theory suggests that the opportunity of crime is determined by potential offender, suitable target, and guardianship (Yar 2005). Note that the suitable target is PII that can reside in physical or digital media and can be held by individuals and organizations. Our goal is to test our hypotheses on how digital and physical channels affect the probability of identity theft victimization through varying proximity to identity. We also control for other factors about potential offender, suitable target and individual

guardianship that are shown in previous studies to have impact on probability of victimization. Below we describe the specific measures for the variables. The detailed operationalizations are shown in Table 8 to Table 10.

4.4.2 Hypothesized Independent Variables

Individual online shopping behavior. We use individual's online shopping behavior to capture the motivated offender's digital channel to compromise PII through the individual. In doing online transactions, the individual interacts with the service application and provide required PII. The PII is then transmitted electronically to the merchant. Thus, a motivated offenders can use phishing attacks and malware attacks to intercept the PII (Choo 2011). Online shopping activity has also been used in previous IDT research to measure online routine activities (Reyns and Henson 2016). Consistent with previous research, individual online shopping is operationalized as a binary variable indicating if an individual has purchased something online. We also include two binary variables that reveal the payment methods used for online shopping, namely whether an individual use only credit card for online shopping, and whether an individual use only debit card for online shopping.

Individual physical theft. Previous research on physical theft shows targets in personal theft include purses and wallets (Mustaine and Tewksbury 1998). Although the thief may intend to steal cash or other valuable objects, PII can also be stolen at the same time. For example, theft of purses and wallets may contain a wide range of PII attributes, such as credit card, debit card, checks and driver's license. Home and car burglars may take away identity documents such as social security cards, medical insurance cards, and tax forms. Therefore, we operationalize physical theft as a binary variable indicating if an individual experienced any events of personal theft, home burglary or car burglary.

Organizational data breach measures whether an individual's PII held in organizational databases is compromised. Since our paper studies at the individual level and does not directly measure organizational characteristics, we operationalize organizational data breach using the items asking whether a respondent has received any data breach notification from organizations. We distinguish whether SSN is contained in PII lost in data breach since SSN is viewed as the most sensitive PII attribute (Berghel 2000). Specifically, we define two binary variables, where data breach without SSN measures whether an individual received data breach notifications and SSN is not contained in the compromised PII; data breach with SSN measures whether an individual received data breach notifications and SSN is contained in the compromised PII.

4.4.3 Control variables

Our hypothesized variables focus on the proximity between motivated offenders and potential victims' identities. RAT also suggests that target suitability and capable guardianship are the other two elements determining the opportunity of crimes. While the structures of proximity are different between identity theft crimes and traditional physical crimes, target suitability and capable guardianship are reasonably transposable from physical crimes to the identity theft (Yar 2005). Previous research also finds evidence that the effects of target suitability and guardianship on the probability of identity theft victimization are consistent with the prediction of RAT. Therefore, we include variables that are related to these constructs as control variables.

Variables	Operationalization	Mean	SD
Guardianship			
Checking credit card/bank statements	Indicator of whether a respondent checked credit card and bank account statements	0.773	0.419
Checking credit bureau reports	Indicator of whether a respondent checked credit reports from credit bureaus	0.372	0.483
Using security software on computers	Indicator of whether a respondent used security software on computers	0.149	0.356
Changing passwords on financial accounts	Indicator of whether a respondent changed passwords on financial accounts	0.287	0.453
Shredding documents containing PII	Indicator of whether a respondent shredded sensitive physical documents that contain PII	0.714	0.452
Purchasing identity theft protection or monitoring services	Average of two binary indicators of whether a respondent purchased identity theft protection service, and credit monitoring services in respective	0.045	0.174
Previous victimization			
Prior IDT	Indicator of whether a respondent experienced victimization of IDT beyond the past year	0.100	0.299
Demographics			
Education	Indicator of whether a respondent has a college degree	0.418	0.493
Employment	Indicator of whether a respondet has a job in the past 6 months	0.567	0.495
Gender	Indicator of whether a respondent is male	0.454	0.498
Age	Age of a respondent at the last birthdate	52.74	17.09
Black	Indicator of whether a respondent is Black	0.101	0.301
Marital status	Indicator of whether a respondent is married	0.604	0.489
Latino	Indicator of whether a respondent is Latino origin	0.122	0.328

Table 9: Individual control variables operationalization and descriptive statistics.

For physical crimes, the suitability of a target can be measured according to its four constituent properties, namely value, inertia, visibility and accessibility (Yar 2005). When the target is PII, the value of PII is determined by the economic value of the identity of an individual that the PII can be used to verify. Unlike value, the other three properties of inertia, visibility and accessibility developed for physical properties as targets are not applicable to informational targets (Yar 2005). Therefore, we control for economic value of the subjects of PII to account for the impact of target suitability. We use income and wealth measures to capture the economic value of the subjects of PII. NCVS-ITS does not have items measuring individual income or wealth for each respondent. Instead, it

measures household income and wealth of each respondent. Specifically, annual household income is measured by a categorized item with 14 levels. We further operationalize it into an ordered categorical variable with 4 levels, in which the intervals for the first three categories are equal (1 = 0 - $24,999, 2 = $25,000 - $49,999, 3 = $50,000 - $74,999, 4 = $75,000 and more) (Burnes et al. 2020). Measures for wealth include number of vehicles owned by the household, whether the household owns the home, and whether the household runs a home business for each respondent.

Guardianship refers to the capability of persons and objects to prevent crime from occurring (Tseloni et al. 2004). The presence of guardianship effects such prevention by providing deterrents to motivated offenders, or direct intervention (Cohen et al. 1980). Deterrent can be provided by a social guardian. For example, when the guardian is a person monitoring a property, it serves a threat to a motivated offender that someone is looking (Felson 2002). In addition to social guardians, guardianship can be security measures that make it harder for a motivated offender to access the target, such as barriers, locks and alarms (Tseloni et al. 2004). We include measures for both social guardianship and security measures.

First, we capture the security measures that an individual takes to protect PII. The NCVS-ITS data set include seven items about respondents' personal behaviors in protecting their PII. Each item uses one question to measure whether a respondent has taken a corresponding protection action (Yes = 1, No = 0). Among them, two items measure online protections, including using security software and changing passwords (Reyns and Henson 2016). Two items measure purchasing behavior of commercial protection services, including credit monitoring services and identity theft protection services (Burnes et al. 2020). Two items measure respondents own monitoring behavior, including checking credit bureau reports, and checking credit card and bank statements (Lai et al. 2012). The

last item measures physical protection of shredding documents containing PII (Milne 2003). We conduct exploratory factor analysis to determine whether the items should be grouped or be treated as distinct variables. Results from principal component analysis shows that the two protections of purchasing credit monitoring service and purchasing identity theft protection service load onto one component. Therefore, we operationalize these two indicator items into one variable of purchasing protection service by taking their average value. The results also show that the remaining five items load onto independent components. We operationalize each of these five items as a binary variable for the corresponding protection action.

In addition to individual actions of taking security measures, we also capture the social guardianship and security measures presented in an individual's household as shown in Table 10. Social guardianship is measured as the number of household members greater than 12 years (Mustaine and Tewksbury 1998). Household security measures capture physical access controls, which are measured by two binary variables including 1) whether the respondent's household lives in a gated community and 2) whether the respondent's household lives in a home that has restricted access to the outside (Mustaine and Tewksbury 1998).

Variables	Operationalization	Mean	SD
Guardianship			
Gated community	Indicator of whether a respondent's household lived in a gated community	0.072	0.258
Restricted access	Indicator of whether a respondent's household lived in a building that had restricted access to the outside	0.064	0.245
Number of household members	Number of roommates older than 12 years living in the same household with a respondent	2.242	1.056
Value			
Household income	Ordered categories of level of annual household income (1 = 0 - $24,999, 2 = $25,000 - $49,999, 3 = $50,000 - $74,999, 4 = $75,000 and more)	2.765	1.127
Number of vehicles	Number of motor vehicles owned by a household	2.080	1.065
Home owner	Indicator of whether a respondent was the owner of the home	0.782	0.413
Home business	Indicator of whether a respondent operated a business from the home address	0.049	0.216
Social context			
Rural residence	Indicator of whether a respondent's household lived in rural areas	0.194	0.395
Residence length	Number of years that a respondent's household has lived in the current home address	15.169	13.027
Times moved	Number of times that a respondent's household has moved in the past 5 years	0.310	0.778
Student housing	Indicator of whether a respondent was living on campus	0.007	0.086
Regiol level controls			
Internet usage rate	Rate of households with Internet use at home	0.742	0.051
Online shopping rate	Rate of individuals who has purchased online	0.398	0.036
Physical theft rate	Rate of individuals who has experienced victimization of personal theft per 1,000 people	1.852	0.258
Log number of data breach	The logarithm of totol number of organizational data breaches	4.953	0.468

Table 10: Household and Regional Control Variables Operationalization and Descriptive Statistics

Apart from the key elements suggested by RAT, we also control for environmental factors that may affect the opportunity of identity theft crimes. First, we consider the social contexts of neighborhood as suggested by previous research (Copes 1999). Second, we consider the broader characteristics of the geographic region of an individual's residence. In addition, we also control for an individual's past identity theft victimization and the individual's demographic characteristics.

The social contexts of neighborhood include factors of adjacency to public facilities, residence in urban areas and stability of the neighborhood community (Mustaine and Tewksbury 1998; Roncek and Faggiani 1985). Since these factors focus on the neighborhood of a household, we only use household characteristics as corresponding measures. Specifically, the measure for adjacency to public facilities is whether a respondent's household lives on campus (Mustaine and Tewksbury 1998). Rural residence is measured directly as whether a respondent's household lives in rural areas. We also include two measures for stability of the neighborhood community, including 1) number of years a respondent's household has been living in the current community, and 2) the frequency of moving in the past 5 years (Mustaine and Tewksbury 1998).

For characteristics of the geographic location of an individual's residence, the NCVS-ITS only provide information at the regional level (i.e., Northeast, South, Midwest, West) to protect the privacy of respondents. We control for four regional characteristics of individual respondents. First, regional online shopping rate captures the overall level of online shopping. It is operationalized as the ratio of number of ITS respondents who have purchased online in the past year over total number of ITS respondents in a specific region. Second, regional internet penetration rate captures the accessibility to computer and internet in a specific region. It is operationalized as the ratio of people who have access to computer and internet over the total population in a specific region (Goel 2019). Third, regional personal theft rate captures the overall risk of personal theft victimization in a specific region. It is operationalized as the ratio of number of larceny-theft victims to per 1,000 people in a specific region. Finally, regional data breach captures the overall risk of organizational data breach in a specific region. It is operationalized as the logarithm of total organizational data breaches in a specific region.

For an individual's past identity theft victimization, previous research shows that past victimization can decrease the probability of future victimization due to negative state dependence (Clay-Warner et al. 2016). This phenomenon can be explained by the "once bitten, twice shy" approach. Experiencing victimization makes people aware of risk, thus motivating them to take self-protective action to prevent future victimization (Cook 1986; Hindelang 1978). We operationalize past identity theft victimization as a binary variable indicating whether a respondent has experienced any type of identity theft victimization beyond the past year (Burnes et al. 2020).

For individual demographic characteristics, We use a consistent set of variables with previous research that are shown to impact victimization, including gender, age, race, Hispanic origin, marital status, education, and employment (Anderson 2006).

4.5 Estimation Model

Our identification assumption is that confounding covariates are either observed in the data or unobserved and time invariant. This assumption is also referred to as the conditional independence assumption (CIA) in econometrics (Angrist and Pischke 2008). In our research context, possible time invariant unobserved characteristic can be individual's disposition of cautiousness. People with the cautious personality may have lower risk of victimization while they may also shop less online. Repeated observations on individuals in our panel data sample allow us to control for the time invariant unobserved confounding covariates by compare within individual variations. We also control for a rich set of observed covariates in the regression model to improve the plausibility of the CIA.

Specifically, we use the logistic regression with fixed effects to model the probability of each type of identity theft victimization. Equation (1) presents the econometric specification for the logit model with fixed effects:

$$\log\left(\frac{p(V_{it}=1)}{1-p(V_{it}=1)}\right) = \beta_1 OLShop_{it} + \beta_2 OLShopCredit_{it} + \beta_3 OLShopDebit_{it}$$
$$+ \beta_4 PhysTheft_{it} + \beta_5 DataBreach_{it} + \beta_6 DataBreachSSN_{it}$$
$$+\lambda \times X_{it} + \gamma \times Z_{it} + \alpha_i + \omega_t \quad (1)$$

For each type of victimization, we fit the above model to a specific type of victimization status V_{it}. Individual respondents are indexed with *i*, and the t-th time of ITS survey with *t*. The coefficients of interest are $\{\beta_k,\ k = 1, 2, \ldots, 6\}$. X_{it} is a vector of control variables for individual protection behaviors and other individual and household characteristics. Z_{it} is a vector of regional level control variables, including the rate of online shopping, physical crimes, data breach, and internet penetration. α_i is a vector of individual fixed effects that absorbs all time invariant individual characteristics. ω_t is a vector of year fixed effects.

In estimating equation (1), α_i can be accounted for by using dummy variables for each individual *i*, and equation (1) can be estimated using the maximum likelihood estimation (MLE). One concern is that when the number of observations is large, MLE can result in inconsistent estimators, which is known as the incidental parameters problem (Allison 2014). A common method to avoid the incidental parameters problem is estimating a conditional fixed effects model (Chamberlain 1980). It maximizes a conditional log-likelihood function using $\sum_t V_{it}$ as the sufficient statistic for the individual fixed effects. Those respondents who do not experience any victimization events ($\sum_t V_{it} = 0$) are dropped from the analysis. We estimate equation (1) using this conditional maximum likelihood method.

4.6 RESULTS

Estimation results based on equation (1) for each of the five types of identity theft victimization are summarized in Table 11. Column 1 to Column 5 present the coefficient estimates of the corresponding type of identity theft victimization. The results show that the digital and physical mechanisms of each type of identity theft are not all same.

	(1)	(2)	(3)	(4)	(5)
Dependent Variable:	Credit Card Fraud	Debit Card Fraud	Other Existing Account Fraud	New Account Fraud	Other Identity Fraud
Online shopping	1.135***	1.412***	1.691***	1.062***	0.654
	(0.115)	(0.116)	(0.314)	(0.317)	(0.475)
Payment with credit card only	0.357***	-0.967***	-0.623**	-0.174	0.624
	(0.081)	(0.098)	(0.216)	(0.266)	(0.405)
Payment with debit card only	-1.176***	0.191†	-0.223	-0.216	0.241
	(0.176)	(0.102)	(0.277)	(0.356)	(0.508)
Physical personal theft	0.641***	0.616***	1.095***	1.002***	0.488
	(0.111)	(0.123)	(0.248)	(0.291)	(0.438)
Data breach	0.378***	0.278*	0.704**	-0.185	1.008*
	(0.086)	(0.114)	(0.234)	(0.382)	(0.391)
Data breach containing SSN	-0.027	-0.151	0.082	1.498***	1.115*
	(0.151)	(0.207)	(0.368)	(0.457)	(0.460)
Controls:					
Individual characteristics	Y	Y	Y	Y	Y
Household characteristics	Y	Y	Y	Y	Y
Regional characteristics	Y	Y	Y	Y	Y
Individual fixed effects	Y	Y	Y	Y	Y
Year-quarter fixed effetcs	Y	Y	Y	Y	Y
Victimization event sizes	1,115	774	131	108	62
Total number of observations	23,376	23,376	23,376	23,376	23,376
Log likelihood	-11189.3	-7773.09	-1317.43	-1086.18	-623.6059

***p<0.001 **p<0.01 *p<0.05 †p<0.1

Table 11: Fixed effects estimates of conditional logit model for identity frauds and IDT victimization.

Based on the results, we explain how they support our theoretical hypotheses. Conclusions on our hypotheses are summarized in Table 12.

	Individual		Organizational
	Digital	Physical	Digital
	Online shopping activity	Personal theft	Data breach
Credit card fraud	H1a(+)	H2a(+)	H3a(+)
Debit card fraud	H1b(+)	H2b(+)	H3b(+)
Other existing account fraud	H1c(+)	H2c(+)	H3c(+)
New account fraud		H2d(+)	H3d(n.s.)
Other identity fraud		H2e(n.s.)	H3e(+)

Note: (+) corresponding coefficient is positive and significant; (n.s.) corresponding coefficient is not significant.

Table 12: Results of hypotheses testing based on regression analysis.

4.6.1 Individual digital mechanisms of identity theft: online shopping activity

The coefficient estimate of individual online shopping activity shows how it is associated with the average probability of identity theft victimization. Results in Table 11, Column 1 to Column 3 shows that purchasing online is positively associated with the three types of fraud that requires low recognitive proximity.

Specifically, Column 1 shows that online shopping increases the log odds ratio of credit card fraud by 1.135 (s.e. = 0.115, $p<0.001$) and H1a is supported. The coefficient estimates on payment method used in online shopping provide further evidence that it is the symbolic representation of credit card that affect the probability of recognitive

proximity and hence influence the probability of credit card fraud victimization. Using only credit card for online shopping increases the marginal log odds ratio of credit card fraud by 0.357 (s.e. = 0.081, $p<0.001$). However, using only debit card for online shopping decreases the marginal log odds ratio of credit card fraud by 1.176 (s.e. = 0.176, $p<0.001$). It supports the idea that the probability of victimization is not increased when the scope of PII used for online shopping does not scope of PII required for credit card fraud. Similarly, Column 2 shows that online shopping increases the log odds ratio of debit card fraud by 1.412 (s.e. = 0.116, $p<0.001$) and H1b is supported. Using only debit card for online shopping increases the marginal log odds ratio of debit card fraud by 0.191 (s.e. = 0.102, $p<0.1$), while using only credit card for online shopping decreases the marginal log odds ratio by 0.967 (s.e. = 0.098, $p<0.001$).

Column 3 shows that online shopping increases the log odds ratio of other existing account fraud by 1.691 (s.e. = 0.314, $p<0.001$) and H1c is supported. Other existing accounts include various types of service accounts such as utility account, mobile phone number account, online payment account. Although we do not have measures on whether these accounts' information is directly used in online shopping activities, the estimates on using only credit card (-0.623, s.e.=0.216, $p<0.01$) and using only debit card (-0.223, s.e. = 0.277) for online shopping are both negative, indicating a substitution effect.

Column 4 shows that online shopping increases the log odds ratio of new account fraud by 1.062 (s.e. = 0.317, $p<0.001$). We expect that only low recognitive proximity is created in online shopping activities, this result suggests that opening new account may involve low recognitive proximity through the online channel. Indeed, some credit card companies offer online application for new credit card account without in-person human recognition. Column 5 shows that the association between online shopping and identity

theft requiring high recognitive proximity is not statistically significant, which is consistent with our theoretical expectation.

4.6.2 Individual physical mechanisms of identity theft: personal theft

The coefficient estimate of physical personal theft shows its association with the average probability of each type of identity theft victimization through Column 1 to Column 5. The results indicate that physical personal theft increases the probability of types of identity theft that require low recognitive proximity. It also increases the probability of new account fraud that requires medium recognitive proximity. However, we do not find an effect on identity theft that require high recognitive proximity.

Specifically, experiencing physical personal theft increases the log odds ratio by 0.641 (s.e. = 0.111, $p<0.001$) for credit card fraud, 0.616 (s.e. = 0.123, $p<0.001$) for debit card fraud, 1.095 (s.e. = 0.248, $p<0.001$) for other existing account fraud, and 1.002 (s.e. = 0.291, $p<0.001$) for new account fraud. For other identity fraud, it does not have a statistically significant effect. It might be difficult for ordinary thief to exploit personal information in all identity dimensions and impersonate as the victim's identity during in-person human recognition. Therefore, hypotheses H2a through H2d are supported, while H2e is not supported.

4.6.3 Organizational digital mechanisms of identity theft: data breach

The coefficient estimate of organizational data breach shows its association with the average probability of each type of identity theft victimization through Column 1 to Column 5. The results provide evidence that organization data breach is positively associated with all types of identity theft victimization, supporting our argument that

personal information in all identity dimensions is possible to be compromised through organizational digital channel.

Specifically, experiencing organizational data breach increases the log odds ratio by 0.378 (s.e. = 0.086, $p<0.001$) for credit card fraud, 0.278 (s.e. = 0.114, $p<0.05$) for debit card fraud, 0.704 (s.e. = 0.234, $p<0.01$) for other existing account fraud, and 1.008 (s.e. = 0.391, $p<0.05$) for other identity fraud. Therefore, H3a, H3b, H3c, and H3e are supported. For new account fraud, data breach does not have a statistically significant effect and H3d is not supported. However, containing SNN in a data breach increases the marginal log odds ratio of new account fraud by 1.498 (s.e. = 0.457, $p<0.001$). This finding can be explained by the requirement of SSN in providing new credit service as stated in the Customer Identification Program (FederalRegister 2003). Similarly, containing SSN in a data breach increases the marginal log odds ratio of other identity fraud by 1.115 (s.e. = 0.460, $p<0.05$).

4.7 ROBUSTNESS CHECKS

In addition to the panel data sample used to estimate equation (1), we also create a matched sample from the pooled cross-section data using the NCVS-ITS in 2012, 2014 and 2016. There are two assumptions used for the cross-section matched sample analysis. First, observations assumed to be independent draws from the population. Second, the values of the independent variables of interest are assumed to be as good as randomly assigned after we match on observable covariates.

Regression analysis based on the matched sample complements the fixed effects model using panel data sample in several ways. First, the fixed effects model estimates the parameters by using the within individual variations. If there is little or no variation in a variable of interest within one individual in different time periods, the fixed effects model

cannot utilize this information for statistical inference. A matched sample looks at the between subject variation for each variable in the regression model. Thus, it can utilize information from a larger number of observations in the NCVS-ITS data. Second, the event size for new account fraud and other identity fraud are not large as shown in Table 11. Using a matched sample with a larger number of observations can also increase the event size used in regression analysis. Third, matching on observable covariates before regression analysis can improve balance in the covariates and reduce model dependence of the results (Ho et al. 2007). We use the coarsened exact matching (CEM) method to create our matched sample, as CEM is considered to be advantageous in improving overall balance in covariates comparing to traditional matching method such as propensity score matching (Iacus et al. 2012).

	(1)	(2)	(3)	(4)	(5)
Dependent Variable:	Credit Card Fraud	Debit Card Fraud	Other Existing Account Fraud	New Account Fraud	Other Identity Fraud
Online shopping	1.171***	1.520***	1.609***	0.564**	0.071
	(0.079)	(0.072)	(0.196)	(0.218)	(0.244)
Payment with credit card only	0.369***	-1.044***	-0.372*	0.429*	0.015
	(0.065)	(0.069)	(0.154)	(0.190)	(0.245)
Payment with debit card only	-1.220***	0.092	-0.498*	-0.233	0.164
	(0.132)	(0.070)	(0.205)	(0.274)	(0.293)
Physical personal theft	0.591***	0.915***	0.757***	0.988***	0.507†
	(0.103)	(0.088)	(0.211)	(0.214)	(0.298)
Data breach	0.272**	0.514***	1.245***	0.466†	1.638***
	(0.096)	(0.098)	(0.187)	(0.280)	(0.250)
Data breach containing SSN	0.547***	0.04	0.028	0.859**	0.699*
	(0.127)	(0.142)	(0.254)	(0.332)	(0.290)
Constant	-4.202***	-3.444***	-5.545***	-4.438***	-4.982***
	(0.113)	(0.109)	(0.273)	(0.274)	(0.329)
Matching covariates					
Individual characteristics	Y	Y	Y	Y	Y
Household characteristics	Y	Y	Y	Y	Y
Regional characteristics	Y	Y	Y	Y	Y
Victimization event sizes	6,168	5,083	917	675	507
Total number of observations	50,975	50,975	50,975	50,975	50,975

***p<0.001 **p<0.01 *p<0.05 †p<0.1

Table 13: Estimates of logistic regression for identity frauds victimization with CEM

Table 13 summarizes the results of logistic regression using the cross-section matched sample. The results are largely consistent with the fixed effects models using

panel data sample. In addition, we also find that organizational data breach has a marginally significant association with new account fraud, which increase the log odds ratio of new account fraud by 0.466 (s.e. = 0.280, $p<0.1$). Physical personal theft also has a marginally significant association with other identity fraud, which increase the log odds ratio of other identity fraud by 0.507 (s.e. = 0.298, $p<0.1$). This evidence provides further support for H2e and H3d, which are not supported by the fixed effects models using the panel data sample. Using the usual statistical significance level of 0.05, we can draw the same conclusions on our hypotheses as that suggested by the fixed effects models using the panel data sample. In conclusion, regression results from the matched sample are consistent with results from the fixed effects models using panel data sample.

Chapter 5: Discussion and Conclusions

This study develops a new concept of recognitive proximity and modifies the existing RAT to explain the antecedents of identity theft victimization. The findings indicate that three channels of individual PII leakage, including individual online shopping activity, individual physical theft, and organizational data breach, creates different levels of recognitive proximity between potential offenders and individual's identity, thus affect different types of identity theft victimization. We first summarize these differences and then discuss the implications of the findings for research and practice.

Individual digital channel through online shopping creates low recognitive proximity. The scope of personal information used in individual online shopping is mainly in the identity dimension of symbolic representation. The results support the hypotheses that individual online shopping activity increases the probability of frauds that require low recognitive proximity, including credit and debit card fraud and other existing account fraud. In addition, the results also show that individual online shopping activity increases

the probability of new account fraud. For opening new accounts in the offline environment, a medium level of recognitive proximity is required for a criminal to successfully impersonate an individual. However, there are credit card companies offering online application of opening new credit card accounts. This practice may reduce that bar for recognitive proximity required for opening new accounts. For identity theft that require a high level of recognitive proximity, online shopping activity does not have an impact on the probability of victimization. It also supports our hypothesis that individual online shopping activity only facilitates the type of frauds that require low recognitive proximity.

Individual physical channel through personal theft creates low to medium recognitive proximity. Although the individual may hold physical documents that contain personal information in all identity dimensions, the results show that individual physical theft increase the probability of types of frauds that involve low to medium level of recognitive proximity. In other words, individual physical theft increases the probability of credit and debit card fraud, and other existing account fraud that require low recognitive proximity, and new account fraud that require medium recognitive proximity. However, it has no impact on the probability of identity theft victimization that require high recognitive proximity. Based on our theorization, high recognitive proximity involves human recognition and the criminal may need some con man ability to behave and communicate like the impersonated identity as in Abagnale's case (Abagnale 2013). Our explanation is that the criminals who conduct physical theft may not have the con man ability to impersonate the victim's identity at high level of recognitive proximity.

Organizational data breach creates the full spectrum of recognitive proximity. Since organizations are collecting increase an increasing amount of data about individuals to facilitate data analytics and business decisions, it also increases the risk of leaking detailed informational portraits about the individuals that involve all identity dimensions. The

results strongly support our hypothesis that organizational data breach will increase the probability of all types of frauds and identity theft.

5.1 CONTRIBUTIONS

This study proposes the concept of recognitive proximity as a new lens for examining the mechanisms of identity theft victimization. To date, researchers have been predominantly focusing on the individual digital mechanism of identity theft victimization. The implicit assumption of this research approach is that online proximity facilitates the occurrence of identity theft. By theorizing recognitive proximity and modifying existing RAT with this approach, this study integrates the individual digital, physical and organizational mechanisms of identity theft victimization.

Specifically, this study argues that online proximity created by individual online shopping activity and physical proximity created by individual personal theft are both related to recognitive proximity. Besides the individual channels of PII leakage, organizational data breach can also create recognitive proximity. In addition, this study provides insights about how the three channels of PII leakage lead to different levels of recognitive proximity. The findings show that individual online shopping activity and individual personal theft both lead to low to medium levels of recognitive proximity. Low to medium levels of recognitive proximity allows the criminals to impersonate the victim's identities and commit credit and debit card frauds, other existing account fraud and new account fraud. This result provides insight that individual physical channel of identity theft is as important as the individual digital channel. It adds to the existing knowledge that identity theft are mainly caused by individual online shopping activity (Reyns and Henson 2016). In addition, the findings show that organizational data breach leads to full spectrum of recognitive proximity and can increase the probability of all types of frauds and identity

theft. It suggests that future research on the mechanisms of identity theft may want to consider the ecosystem of identity theft including both individual and organizational channels.

The theoretical development of RAT using the recognitive proximity approach also provides a framework to unify the definition of identity theft. The lack of commonly agreed definition of identity theft has raised challenges for identity theft research (Cheney 2005). One reason leading to this difficulty is that existing statutory definition of identity theft takes a reductive approach and defines the misuse of any type of PII attributes as identity theft. This definition leads to ambiguity in deciding which type of frauds should be categorized as identity theft, since all types of frauds involves misuse of certain scope of PII. We argue that identity theft requires impersonating the victim's identities by exploiting the stolen PII, not just leakage and misuse of PII. We provide a clear scale of recognitive proximity, which can be used to map the level of impersonation required for each type of fraud. Categorization of identity theft can be done with clarity by distinguishing the level of recognitive proximity involved in each type of fraud. The findings support the idea that the mechanisms of identity theft victimization are crime specific, and it is desirable to distinguish different types of identity theft victimization.

5.2 LIMITATIONS

This study utilizes the NCVS-ITS survey data to test the hypothesized digital and physical mechanisms of identity theft victimization. The measurement items of the government's ITS survey are consistent with the measures used in previous research and the survey is carefully designed to minimize recall bias. However, it is subject to the general limitations of an archival survey data set. First, the victimizations are measured based on survey respondent's report instead of actual crime events. Second, the causal

interpretation of the estimation results depends on the conditional independence assumption. Third, the measurements for the five categories of identity theft victimizations may not be sufficient to capture the evolvement of new identity theft forms which may have developed during the post-pandemic digital economy. Future research is needed to address these limitations.

DATA BREACHES AND IDENTITY FRAUD RISKS OF INDIVIDUALS

Abstract

We develop a theory to explain whether, how, and when individuals can protect themselves against the heightened identity theft (IDT) risks following a data breach. We conceptualize IDT as a multi-stage process where criminals first unlawfully obtain a person's identifying information (PII) through a data breach, then misuse the PII to assume the identity of the person, and ultimately, imposter as the person to commit the IDT crime. We distinguish if the person's PII leaks to criminals through a personal data breach or an organizational data breach. We hypothesize that preventive protections can reduce the PII leakage through personal data breaches but they can increase the PII leakage through organizational data breaches. We further hypothesize that detective protections can mitigate the heightened IDT risks following a data breach. Finally, we hypothesize that proactive protections are likely to be more effective than reactive protections in reducing IDT. We find support for these ideas in a representative sample of citizens in the U.S.A.

Chapter 6: Introduction

Data breaches leak personal identifying information (PII) to unauthorized parties. In 2019, there was a 17% increase in data breaches over the previous year (Wysopal 2020). An implication of the data breaches for individual victims whose PII leak is the heightened risk of identity theft (IDT). IDT is the fastest growing crime in the U.S.A (van der Meulen 2011). In 2016, an estimated 26 million people were victims of the IDT crime (Harrell 2019). The total economic loss of $17.5 billion due to IDT crimes (Harrell 2019) exceeded the total economic loss of $16.4 billion due to property crimes (FBI 2019). IDT victims are responsible for dealing with creditors and debt collectors to clear their names and credit files (LaPiedra 2014). Beyond financial losses, IDT victims experience emotional symptoms such as distress and anxiety and physical symptoms such as headaches and trouble sleeping (Golladay and Holtfreter 2017). IDT also has long-term effects on victims: e.g., inability to trust people and a belief to have lost everything (van der Meulen 2011). People who worry about becoming an IDT victim also tend to refrain from conducting digital transactions (Anderson et al. 2008; Hille et al. 2015; Pratt et al. 2010). Given the severity of these consequences for both individuals and the digital economy, it is important to understand which protections individuals can take before or after a data breach to reduce their IDT risks.

As summarized in Table 14, there is a growing literature on IDT. However, the sub stream of the IDT literature focusing on the effectiveness of individual protections against IDT is in its infancy. Of the 50 papers we identified on IDT, only 13 were related to individual protections, and they all focused on the individuals' adoption of protections (Jansen and Van Schaik 2018; Kim and Kim 2016; Li et al. 2019; Ogbanufe and Pavur 2016). Only three of the 13 studies examined if the adopted protections were in turn effective in reducing the IDT risk (Burnes et al. 2020; Lai et al. 2012; Reyns and Henson

2016). One study found that individual protections reduced the likelihood of IDT (Lai et al. 2012). Another study found that individual protections did not reduce the likelihood of IDT (Reyns and Henson 2016). A third study found that one of the most commonly recommended protections, the use of identity theft monitoring and protection services, increased IDT rather than decreasing it (Burnes et al. 2020). Given the mixed state of these findings, we develop and test a new theory to explain what types of individual protections are effective in reducing IDT risk.

Research Focus		Summary Findings	Selective Sources
Conceptualization of IDT	Definition of IDT	IDT definitions have been inconsistent. Some studies focus on misuse of another person's PII; some focus on unlawful collection and misuse of another person's PII; some focus on appropriation of another person's identity with intent to commit fraud or other crimes.	Bose and Leung 2019; Wang et al 2017; Lai et al 2012.
	Types of IDT	Five types of IDT have been studied: 1) existing credit card fraud; 2) existing bank account fraud; 3) other existing account fraud; 4) new account fraud; 5) other fraudulent activities.	Burnes et al 2020; Kalvet et al 2019; Lai et al 2012; Anderson 2006.
	Stages of IDT	Most studies treat IDT as a single-stage event. Some studies conceptualize IDT as a multi-stage process where a criminal: 1) unlawfully obtains PII of a person; 2) misuses the PII to steal the identity of the person, and 3) imposters as the person to commit a fraud, theft, or other unlawful activity using the person's identity.	Albrecht et al 2011; Newman and McNally 2005.
Antecedents of IDT	Demographics	Demographic characteristics have been found to impact the likelihood of IDT, including age, gender, race and ethnicity, education, and income.	Copes et al 2010; Anderson 2006.
	Risk factors	Other risk factors of IDT that have been studied include online shopping behavior, data breach experience, previous IDT victimization, and victimization of other crimes.	Goel 2019; Navarro and Higgins 2017;
Protections against IDT	Types of protection	Some common types of individual protections have been studied: e.g., use antivirus software, change passwords regularly, shred documents, check credit card/bank statements, review credit reports, purchase IDT monitoring and protection services.	Albrecht et al 2011; Milne 2003.
	Individual adoption of protection	Of the 50 IDT papers reviewed, 13 study individual protections against IDT. Majority of them focus on adoption of protections. They use a variety of theories to conceptualize the protections: e.g., protection motivation theory, threat avoidance theory, routine activity theory, etc.	Ylang 2020; Jansen and Schaik 2018; Ogbanufe and Pavur 2016; Kim and Kim 2016.
	Effectiveness of protection	There is a dearth of research on the effectiveness of IDT protections. Only three papers have empirically examined whether protections reduce IDT risk. Their results are mixed and inconclusive.	Burnes et al 2020; Reyns and Henson 2016; Lai et al 2012.
Consequences of IDT	Financial impact	In 2016, 26 million persons over the age of 16 were victims of IDT. Total economic loss associated with IDT was about $17.5 billion.	Harrell 2019.
	Health impact	IDT victims also experience emotional (e.g. depression) and physical symptoms such as headaches, high blood pressure, trouble sleeping, upset stomach, and fatigue.	Golladay and Holtfreter 2017.
	Behavioral changes after IDT	After the IDT victimization, victims tend to change their credit behavior, payment choices, and adoption of IDT protection services such as use of IDT protection services, credit alerts, and freezes on credit files.	Li et al 2019; Kahn and Linares-Zegarra 2016; Reyns and Randa 2017.

Table 14: Review of identity theft literature.

To commit the IDT crime, criminals go through a multiple-stage process. First, they gain unauthorized access to some elements of a person's PII through a data breach. Then, they use the compromised PII to gain access to more elements of the person's PII. They try to accumulate sufficient PII to be able to imposter as the person and conduct transactions. Finally, they exploit the stolen identity to imposter as the person and commit fraud, theft, or other crimes (Albrecht et al. 2011; Newman and McNally 2005a). Prior empirical studies focused only on the last stage of this process. They did not theorize which protections might be effective in the earlier stages of the IDT process. We hypothesize that preventive protections can be effective in preventing the leakage of a person's PII to criminals whereas detective protections can be effective in mitigating the IDT risk after the PII leaks. We distinguish if PII leaks to criminals through a personal data breach or an organizational data breach. We hypothesize that individual protections can be effective in reducing PII leakage through individual data breaches, but counterintuitively, they can increase PII leakage through organizational data breaches. Finally, previous research recognizes that an individual's adoption of protections is correlated with the individual's characteristics and risk profile (Ylang 2020). Yet, the extant IDT studies do not take into account the endogenous adoption of the protections. To be able to study the causal effects of the protections on IDT, we use propensity score matching to account for potential bias due to observable characteristics and risk profiles of individuals. We find support for the proposed theory in a representative sample of citizens in the U.S.A.

Chapter 7: Theoretical Foundations

7.1 Conceptualization of IDT

Identity. The concept of identity is defined as "the distinguishing character or personality of an individual" (Merriam-Webster 2020).

Authentication of Identity. A key problem is to authenticate the identity of a person without causing too much inconvenience. Historically, three factors have been used for authenticating the identity of a person in digital transactions (Steinbart et al. 2016): (a) what you know (e.g., username, password, PIN, answers to questions like 'what is your pet's name?'); (b) what you have (e.g., a smart card, a USB token); and (c) what you are (e.g., biometric identifier such as fingerprint). Recently, two additional authentication factors are emerging (Young 2020): (d) what you do (e.g., behavioral biometrics such as how you type); and (e) where you are (e.g., geolocation).

IDT Definition. IDT is defined as the appropriation of another person's identity to commit illegal activities such as fraud, theft, and other crimes (Bose and Leung 2013; Bose and Leung 2019; Milne 2003).

IDT Process. A criminal has to go through a multi-stage process to commit the IDT crime (Burnes et al. 2020; Cheney et al. 2014; Copes et al. 2010; Eisenstein 2008; Kahn and Roberds 2008; Reyns and Henson 2016; Wang et al. 2017). First, the criminal has to gain nonconsensual, unauthorized access to some elements of a person's PII through a data breach. The criminal can access the PII through a digital breach of the person's private computing devices, or a manual breach of the person's physical media such as paper files. The criminal can also access the PII through a breach of an organization that holds the person's PII. Usually, the PII that leaks in a data breach is limited in scope. It may not be sufficient to steal the identity of the person. For example, if a data breach leaks only the person's username, password, and mailing address, the criminal cannot use such PII alone to open a new credit card or bank account in the name of the person. The criminal may also need to gain access to additional elements of the person's PII such as date of birth, social security number (SSN), driver's license, passport, etc. In the second stage, the criminal can search for additional compromised PII of the person or attempt to take over additional

accounts of the person in order to accumulate enough PII and assume the identity of the person. In the third stage, the criminal misuses the accumulated PII to imposter as the person. Finally, the criminal commits an illegal activity such as fraud, theft, and other crime in the name of the person by exploiting the accumulated PII of the person (Albrecht et al. 2011; Newman and McNally 2005a).

IDT Types. Prior research identified the following types of illegal activities as the most common types of IDT crime: (1) existing credit card fraud, (2) existing bank account fraud, (3) other existing account fraud (e.g., fraud in wireless phone account, utility account, insurance, etc.), (4) new account fraud (e.g., open new credit card account, new bank account, etc.), and (5) other fraudulent activities such as applying for a job or government benefits, and filing false tax returns using another person's identity (Burnes et al. 2020; Copes et al. 2010; Kalvet et al. 2019; Lai et al. 2012).

7.2 Conceptualization of Individual Protections Against IDT

Academic studies on individual protections against IDT risk focus on six major types of protections, as summarized in Table 15: (i) securing personal computing devices; (ii) changing account passwords frequently; (iii) shredding personal sensitive documents; (iv) purchasing IDT monitoring and protection services; (v) checking credit card and bank statements regularly; and (vi) checking credit bureau reports.

Type of Protection	Specific Measurement Items	Concepts	Underlying Theories	Selective Sources
Securing computing devices	Use computer security software	Guardianship	L-RAT	Burnes et al 2020
	Use anti-virus software	Online self-guardianship	RAT	Reyns and Henson 2016
	Use computer security software	Capable guardianship		Ylang 2020
	Use anti-virus and firewall tools; Frequently update computer system	Coping behavior	TTAT	Lai et al 2012
	Use firewall and anti-virus software	Countermeasures, safeguards, and controls	IS Control Theory; Information Security Action Cycle Theory	Eisenstein 2008; Straub 1989; Straub and Welke 1998
	Use antivirus software	IDT prevention	No Theory	Gilbert and Archer 2012
Changing passwords	Change passwords on financial accounts	Guardianship	L-RAT	Burnes et al 2020
	Change passwords on financial accounts	Capable guardianship	RAT	Ylang 2020
	Change passwords	Online self-guardianship		Reyns and Henson 2016
	Use of password access control	Preventive protection	Information Security Action Cycle Theory	Straub 1989; Straub and Welke 1998
Shredding sensitive documents	Shred documents	Guardianship	L-RAT	Burnes et al 2020
	Shred documents	Capable guardianship	RAT	Ylang 2020
	Shred bank and credit card statements	Coping behavior	TTAT	Lai et al 2012
	Shred documents	Countermeasures, safeguards, and controls	IS Control Theory; Information Security Action Cycle Theory	Eisenstein 2008; Cram et al 2016b; Straub and Welke 1998
	Shred documents	IDT prevention	No Theory	Gilbert and Archer 2012
Purchasing IDT protection services	Purchase IDT protection service	Guardianship	L-RAT	Burnes et al 2020
	Purchase IDT protection service	Capable guardianship	RAT	Ylang 2020
	Use of IDT protection service	Protection behavior	PMT	Kim and Kim 2016
	Use of credit monitoring service			Ogbanufe and Pavur 2016
	Use of credit monitoring service	Countermeasures, safeguards, and controls	IS Control Theory; Information Security Action Cycle Theory	Eisenstein 2008; Cram et al 2016b; Straub and Welke 1998
Checking credit card and bank statements	Check bank and credit card statements	Guardianship	L-RAT	Burnes et al 2020
	Check bank and credit card statements	Capable guardianship	RAT	Ylang 2020
	Review bank and credit card statements	Coping behavior	TTAT	Lai et al 2012
	Monitor bank and credit card	IDT detection	No Theory	Gilbert and Archer 2012
Checking credit bureau reports	Check credit report	Guardianship	L-RAT	Burnes et al 2020
	Check credit report	Capable guardianship	RAT	Ylang 2020
	Monitor credit records	Coping behavior	TTAT	Lai et al 2012
	Get credit report	IDT detection	No Theory	Gilbert and Archer 2012
	Use of fraud alerts and credit freezes	Consumer protections against IDT	No Theory	Cheney et al 2014

Table 15: Measures, constructs, and underlying theories for individual protections against IDT risk.

Different studies use different theories to conceptualize a given type of protection, as summarized in column 4 of Table 15: e.g., protection motivation theory (PMT) (Kim and Kim 2016; Ogbanufe and Pavur 2016); routine activity theory (RAT) or lifestyle routine activity theory (L-RAT) (Burnes et al. 2020; Reyns and Henson 2016); technology

threat avoidance theory (TTAT) (Lai et al. 2012); IS control theory (Cram et al. 2016a); information security action cycle theory (Straub 1989; Straub and Welke 1998).

To identify which of these theories would be the most appropriate for our study, we use two criteria: (a) theory must focus on the effectiveness of protections; and (b) theory must discuss which protection might be effective in which stage of the IDT process. Theories such as the PMT focus predominantly on the conditions under which individuals adopt protections. They assume that the protections would subsequently be effective in reducing the risk, if adopted and used as prescribed. Theories such as RAT and L-RAT focus on the conditions under which a crime would take place: e.g., a likely offender, a suitable target, and the absence of a capable guardian would create the conditions of a crime. Theories such as the TTAT link the protections directly to IDT. They do not explain which protections would be effective in which stages of the IDT process. Only the information security action cycle theory satisfies both of our criteria simultaneously. Thus, we build on the information security action cycle theory. We also note that all of these theories are consistent with each other regarding the role of a protection. Namely, the objective of a protection is to serve as an individual control mechanism for reducing the risks of a data breach or IDT by restricting access to the individual's PII and making it more difficult, costly, and risky for an unauthorized party to gain access to the PII. We use this common theoretical logic in developing our hypotheses.

The information security action cycle contains four stages and the corresponding types of protections: (1) deterrence, (2) prevention, (3) detection, and (4) remedies (Straub and Welke 1998). Deterrence stage recommends the use of protection mechanisms that provide disincentives to potential criminals by increasing the certainty of sanctioning and the severity of sanctioning (Blumstein et al. 1978). Deterrence changes criminals' cost-benefit calculus (Becker 1968) so that they would not consider attempting an attack. In the

IDT context, government makes the severity of sanctioning visible to potential identity thieves by establishing laws and regulations on IDT crimes (The United States Department of Justice 2017; US Public Law 1998; US Public Law 2004). Such deterrence protections are beyond the control of individuals. Thus, we do not include deterrence as part of individual-level protections against IDT. However, we recognize that preventive protections of individuals could play a deterrence role by making it more difficult for criminals to access the PII residing on personal computing devices and files. We also leave the remedy stage of the information security action cycle out of scope for this study because it focuses on remedying the harmful effects of IDT once the IDT occurs and causes harms (Straub and Welke 1998). Our goal in this paper is to identify the protections that prevent the IDT in the first place, or detect the IDT soon after it happens. Thus, we focus on the prevention and the detection stages of the information security action cycle.

Preventive protections are countermeasures, safeguards, or controls that are designed to prevent the occurrence of threat events such as PII leakage (Cram et al. 2016b; Virtue and Rainey 2015). Preventive protections aim to deprive criminals of opportunities, power, or hope of conducting data breach attacks and succeeding. They have capabilities to ward off illegitimate access of use (Gopal and Sanders 1992; Gopal and Sanders 1997). Preventive protections can make it more difficult and costly for criminals to gain unauthorized access to a person's PII and stop the criminals from progressing to the later stages of the IDT process. The diminished ability to gain access to the person's PII through the person's computing devices and storage media can motivate the criminals to gain access to the same PII through alternative channels. Typically, individuals share their PII with organizations, and organizations store the PII in their databases (Anderson et al. 2008). Criminals can attempt to breach the organizations to gain access to individuals' PII.

Detective protections are controls designed to discover if a threat event of interest such as a PII leakage event or an IDT event actually occurred (Cram et al. 2016b; Virtue and Rainey 2015). Detective protections gather evidence of misuse of a person's stolen PII or identity (Straub and Welke 1998). Thus, detective protections can have two effects in the IDT process: (a) mitigation effect, or (b) discovery effect. As shown in Figure 3.1a, when a person's PII leaks to criminals through a data breach at time t_2, there is a time lag before the criminals can misuse the leaked PII to steal the identity of the person and commit the IDT crime at time t_4. In this period, the criminals can attempt to misuse the leaked PII or supplement it with additional PII so that they can steal the identity of the person. Detective protections can potentially discover the *suspicious PII usage activities* (e.g., PII is sold in the dark web or the PII is being used in attempts to breach additional accounts of the victim). By alerting the victim to the misuse of the leaked PII, detective protections can enable the victim to stop the misuse and disrupt the criminals' activities to profit from the leaked PII to cause damages to the victim. We call this the *mitigation effect* as the detective protections can mitigate the heightened IDT risk following the PII leakage in a data breach. Second, if the mitigation effect fails and the criminals succeed in committing the IDT crime, detective protections can potentially discover that the IDT crime occurred. We call this effect the *discovery effect* of detective protections.

We also distinguish if the protections are taken *proactively* well in advance of any PII leakage; or *reactively*, in response to a data breach notification. An organizational level study found that only proactive protections are effective in reducing data security events (Kwon and Johnson 2014). In the IDT context, the industry practice assumes that if an individual adopts protections reactively after a data breach notification, the reactive protections against the IDT risk could be better than no protections at all. Building on these conceptual foundations, we propose the research model in Figure 3.1b.

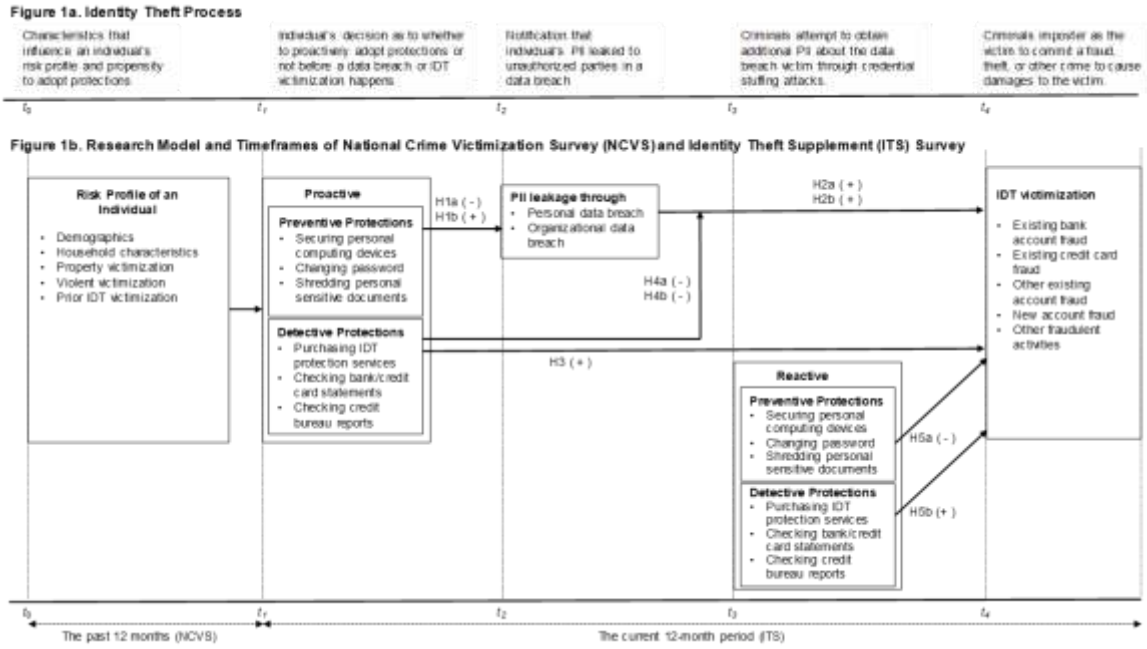

Figure 3: IDT process and research model.

Chapter 8: Hypotheses Development

8.1 Proactive Preventive Protections and PII Leakage

IDT research shows that criminals can attempt to gain unauthorized access to PII through physical means such as stealing postal mail and doing dumpster diving, or digital means such as hacking (Kalvet et al. 2019; Zaiss et al. 2019). The information security action cycle theory suggests that preventive protections can make such attempts more difficult and costly (Gopal and Sanders 1992; Gopal and Sanders 1997). For example, shredding documents is a preventive protection that makes it more difficult and costly for a criminal to obtain PII through dumpster diving. Some examples of preventive protections for securing PII in the digital medium are: securing one's personal computing devices by using anti-virus software, malware detection software, encryption; and using complex

passwords and regularly changing them. Individuals who do such preventive protections can block criminals' access to their PII. Highly motivated and resourceful criminals can still eventually compromise the PII (Moore et al. 2009). However, relative to individuals who do not proactively take preventive protections, individuals who take them are likely to reduce the likelihood of PII leakage through personal data breaches (personal PII leakage).

H1a. *Individual's proactive preventive protections reduce the likelihood of personal PII leakage.*

Government agencies and private organizations also collect and store individuals' PII (Anderson et al. 2008). Individuals do not have full control over the protection of their PII residing in organizational databases. Using individual preventive protections may not be able to reduce the PII leakage through organizational data breaches (organizational PII leakage). It can even increase organizational PII leakage by changing criminals' cost-benefit calculus. Criminals analyze the costs, risks, and benefits of their attacks (Becker 1968). They look for any sources of PII that can be accessed at the lowest cost possible (Anderson et al. 2008). The risk of PII leakage is thus determined by the holder of PII that has the "weakest link" of access control, or the "path of least resistance" (Varian 2004). If obtaining the same PII is easier, less costly, and less risky through an organizational data breach, criminals can target the organization rather than the individual. By securing personal computing devices, changing account passwords frequently, and shredding sensitive paper files, an individual can make it much more difficult and costly for criminals to obtain the PII directly from the individuals' computing devices and paper files. When criminals cannot access the PII through the individual, they are likely to try to gain access to copies of the same PII residing in organizational databases. Thus, individuals who take proactive preventive protections can face higher risk of organizational PII leakage.

H1b. *Individual's proactive preventive protections increase the likelihood of organizational PII leakage.*

8.2 Proactive Detective Protections and IDT Victimization

8.2.1 Discovery Effect of Proactive Detective Protections

As Hoofnagle (Hoofnagle 2005) notes, many victims of IDT do not discover that their identity was stolen and exploited until after negative consequences happen and harms accumulate. Victims typically find out about the IDT occurrence when debt collectors contact them about overdue payments on loans taken out in their names; when fraudulently filed applications for credit, loans, employment, housing, etc. start becoming denied; when law enforcement contacts them; and when their utilities and other accounts become turned off by service providers due to lack of payments (Velasquez et al. 2017). Once criminals take control of a victim's accounts, they change the contact information on the accounts so that the victim cannot receive communications and does not become aware of the fraudulent transactions the criminals make on the accounts (Albrecht et al. 2011). Information security action cycle theory argues that detective protections can potentially discover the misuse of PII and fraudulent activity by gathering evidence of misuse (Straub and Welke 1998). Seeing evidence of misuse is important for discovering that IDT victimization occurred. Among the protections studied in the previous IDT literature (Burnes et al. 2020; Milne 2003; Reyns and Henson 2016), we focus on three types of detective protections: purchasing IDT monitoring and protection services (IDT protection services), checking credit card/bank statements, and checking credit bureau reports. These protections can help victims discover if there are fraudulent transactions on their credit card and bank accounts, and if new accounts have been opened in their names, etc. (Barron and

Staten 2003). Individuals who take these detective protections proactively are more likely to discover IDT victimization compared to individuals who do not take them. Thus:

H3. *Proactive detective protections increase the individual's likelihood of discovering IDT victimization.*

8.2.2 Mitigation Effect of Proactive Detective Protections

Detective protections can also potentially play a mitigation role by negatively moderating the link between PII leakage in a data breach and IDT victimization. As H2 states, when PII leaks in a data breach, IDT risks of data breach victims go up. However, as shown in Figure 1a, there is a time period in between the PII leakage and the IDT victimization stages of the IDT process. If the PII that leaked in the data breach is comprehensive enough to steal the identity of a person, criminals may be able to exploit the PII for fraud immediately. However, the PII that leaks in a data breach is typically limited in scope: e.g., name, address, credit card number; etc. Criminals may have to obtain additional pieces of PII of a data breach victim to be able to steal the victim's identity (e.g., date of birth, SSN, Passport#, Driver's License#, etc.). Criminals use the PII obtained from the initial data breach to try to breach additional accounts of the victim. They write bots to check if the victim was part of any previous data breaches where additional PII about the victim might be available. Criminals do "credential stuffing" attacks by trying the login credentials of the compromised accounts in additional accounts of the same person. Criminals continue such activities until they gather enough PII to assume the identity of the person and imposter as the person in attempting a fraud (Miller-Osborn 2017). Criminals may also have to wait a while for approval after submitting fraudulent applications to open up new accounts, obtain government benefits, tax returns, etc. in the name of the person (Albrecht et al. 2011). Detective protections can potentially enable the

person to detect some of these fraudulent activities. For example, using IDT monitoring and protection services as a detective protection can alert the person if any other accounts are being compromised and if any additional pieces of the person's PII appear on the dark web. Detecting such activity can enable the person to make interventions such as closing some accounts, freezing some accounts, placing fraud alerts, etc. Such interventions could disrupt, delay, and stop the successful completion of the IDT process. Similarly, using detective protections such as reading credit bureau reports and reviewing credit card and bank statements regularly could alert the person to any fraudulent transactions in a timely manner. For example, setting a mobile phone alert on credit card transactions can instantly inform the person about a fraudulent charge attempt, deny it, request an investigation or the cancellation and replacement of the credit card. Such interventions could stop the IDT process before criminals are able to complete the illegal activity and cause any loss to the person. Thus, compared to individuals who do not take proactive detective protections, individuals who take them are likely to mitigate the heightened IDT risk after their PII leak in a data breach.

H4a. *Proactive detective protections negatively moderate the link between personal PII leakage and IDT victimization.*

H4b. *Proactive detective protections negatively moderate the link between organizational PII leakage and IDT victimization.*

8.3 Reactive Protections and IDT Victimization

Data breach notification letters advise victims to take protections against heightened IDT risks following the leakage of PII. If a data breach victim did not take proactive protections, the breach notification letter could convince the victim to take protections reactively.

Taking the preventive protections reactively in response to a data breach notification letter is unlikely to reduce the data breach victim's subsequent IDT victimization. Preventive protections aim to restrict the access of criminals to identity related information (Burnes et al. 2020; Cram et al. 2016b; Eisenstein 2008; Lai et al. 2012; Virtue and Rainey 2015). If some PII already leaked, it becomes easier for criminals to misuse the leaked PII to gain access to more PII of the data breach victim. Criminals can explore if the leaked PII allows access to additional accounts of the victim. They can also install key loggers on the victim's devices. If the victim changes account passwords in reaction to a data breach notification, criminals can gain access to the changed passwords as well and use them to breach more accounts of the victim. The accumulation of PII and compromised accounts increases the likelihood of IDT victimization. Nevertheless, compared to a data breach victim who does not take any protections at all, a data breach victim who starts to secure personal computing devices and shred sensitive documents after receiving a data breach notification letter, is more likely to reduce IDT victimization. Thus:

H5a. *Reactive preventive protections reduce individual's likelihood of IDT victimization.*

Taking the detective protections reactively can increase the data breach victim's likelihood of discovering IDT victimization (Eisenstein 2008). Detective protections aim to discover if threat events such as IDT victimization occur (Cram et al. 2016b; Virtue and Rainey 2015). It is uncertain whether the PII that leaks in a data breach subsequently becomes abused for IDT crimes (Acquisti et al. 2016; Mann 2015). Reactive detective protections, such as starting to read credit card and bank statements, reviewing credit bureau reports, and subscribing to IDT monitoring and protection services, can gather evidence of PII abuse (Straub and Welke 1998). If criminals abuse the leaked PII for

committing an IDT crime, the reactive detective protections can help the data breach victim to discover the IDT victimization. Thus:

H5b. *Reactive detective protections increase individual's likelihood of discovering IDT victimization.*

Chapter 9: Methods

9.1 DATA

Our hypotheses examine how and when individual protections are effective. To test our hypotheses, we use three waves of National Crime Victimization Survey (NCVS) data that contain detailed information about individual protection, PII leakage and identity theft victimization. The NCVS collects annually individual level crime information from a nationally representative panel. Identity theft crime data were collected in 2012, 2014 and 2016 with an Identity Theft Supplement (ITS) in the NCVS. The ITS survey in a given year (t) asked questions about each stage of the IDT process in Figure 1b. Specifically, the ITS asked: (1) if an individual took any proactive protections (t_1) before receiving any notification of a data breach (t_2); (2) whether the individual received notification of a data breach (t_2); (3) whether the individual took any reactive protections (t_3) in response to a data breach notification (t_2); and (4) whether the individual subsequently experienced an IDT victimization (t_4). The NCVS and ITS data have been used in prior academic studies on IDT (Burnes et al. 2020; Golladay and Holtfreter 2017; Navarro and Higgins 2017; Reyns and Randa 2017).

The longitudinal survey of NCVS and ITS allows us to test our hypotheses with both a pooled cross-sectional sample and a panel data sample. The measurements for dependent and hypothesized independent variables are from ITS in 2012, 2014, and 2016. We exclude ITS data in year 2008 because it did not ask questions about the individual

protections against IDT. We first created a panel data sample that track individuals ITS respondents in different years. This study is the first study that provides empirical evidence with longitudinal data from ITS. Previous academic research has used ITS data cross-sectionally because unique respondent tracking IDs across years are not directly provided in the original data. However, we scrutinize the survey design and find it possible to track a portion of respondents for two ITS surveys. The constructed panel dataset allows us to utilize within subject variation to address potential selection problem discussed below.

Nevertheless, we also created a pooled cross-sectional data set using the three waves of ITS data. We paired a respondent's ITS data in year (t) with the respondent's corresponding NCVS data in year (t-1), i.e., 2011, 2013, and 2015, to be able to use the rich set of socio-demographic variables in NCVS for the matching method. This approach utilizes the between subject variation in the following data analysis. It provides complementary evidence to the panel-data analysis in addressing the potential selection problem discussed below. The combined overall NCVS-ITS response rates for 2012, 2014 and 2016 are respectively 68.2%, 66.1% and 60.0%. The BJS also tested whether respondents and non-respondents differ in important ways, but it did not find any nonresponse bias (Burnes et al. 2020).

9.2 MEASURES

9.2.1 Dependent Variables

Table 16 presents measurement instruments used in ITS to measure the dependent variables.

DVs	Measurement Items and Response Scale	Sources
OrgPIILeak	During the past 12 months, has a company, government, or some other organization that has your personal information on file ever notified you that paper or electronic files containing your personal information may have been lost, stolen or posted on a publicly available website? (1=Yes, 0=No)	Burnes et al 2020; Mikhed & Vogan 2018
PsnPIILeak	If respondent reported leakage of PII, but it was not due to an organizational breach (based on the OrgPIILeak question above), personal PII leakage is [1]; otherwise, it is [0].	
IDTVictimization	Instruction: I would like to ask you questions about identity theft. Identity theft means someone else using your PII without your permission to buy something, get cash or services, pay bills or avoid the law. During the past 12 months, has someone, without your permission used or attempted to use your (1=Yes, 0=No): 1) Misuse of bank accounts: Existing checking or saving account, including any debit or ATM cards? 2) Misuse of credit cards: One or more of existing credit cards? Please do not include debit cards. 3) Misuse of other existing accounts: Another type of existing account such as telephone, cable, gas or electric accounts, online payment account insurance policies, entertainment account or something else? 4) New account fraud: Personal information to open any NEW accounts such as wireless telephone accounts, credit card accounts, loans, bank accounts, online payment accounts or something else? 5) Other fraudulent purposes: Personal information for some other fraudulent purpose such as getting medical care, a job, or government benefits; renting an apartment or house; giving your information to the police when they were charged with a crime or traffic violation, or something else? If a respondent reported any one of these crimes, IDT victimization=[1]; otherwise [0].	Burnes et al 2020; Reyns & Henson 2016; Lai et al 2012; Copes et al 2010; Anderson et al 2008; Anderson 2006

Table 16: Measurement instruments, sources, and summary statistics of dependent variables.

PII Leakage. We measure if a respondent's PII leaked to unauthorized parties in either a personal data breach or an organizational data breach. *Organizational PII leakage (OrgPIILeak)* = [1] if individual received a data breach notification letter from an organization in the past 12 months; or [0] otherwise. *Personal PII leakage (PsnPIILeak)* = [1], if individual reported a data breach, but it was not an organizational data breach; or [0] otherwise.

IDT Victimization. Consistent with the academic studies in Table 1, ITS measures five types of IDT crimes in which a person's PII was misused to steal the identity of the person to commit a fraud: (1) misuse of existing credit cards; (2) misuse of existing bank accounts; (3) misuse of other types of existing accounts such as telephone, utility, online payment account, and insurance policy; (4) opening any new accounts; and (5) other frauds such as applying for a job, government benefits, or housing, filing false tax returns, and

avoiding arrest from law enforcement. According to the theoretical framework as discussed in Chapter 2.5, these five frauds are different crimes. Empirical evidence presented in Chapter 4.4.1 also supports the idea of viewing them as different crimes. Therefore, we create a binary variable for individual's victimization of crime *j*, *j* is in [1,2,3,4,5].

9.2.2 Independent Variables

Table 17 presents measurement instruments used in ITS to measure individual protections against IDT. The ITS survey asked a series of questions to capture six IDT related protections of individuals within the past 12 months. We categorized the protections into preventive and detective types of protections using the definitions of these concepts (Cram et al. 2016b; Virtue and Rainey 2015). Specifically, the three preventive protections in ITS are: securing personal computing devices (*SecurePC*), shredding personal sensitive documents *(ShredDoc*), and changing financial account passwords frequently (*ChangePassword*). The three detective protections in ITS are: purchasing IDT monitoring and protection services (*PurchaseIDTprotection*); regularly checking credit card and bank statements (*CheckStatement*); and checking credit bureau reports (*CheckCreditReport*). An affirmative response to each protection question triggered a follow-up question in ITS asking whether the respondent took the said protection in response to a PII leakage in a data breach. We coded the said protection as proactive [1] if it was taken in advance of any PII leakage in a data breach; or [0] otherwise (Burnes et al. 2020). We coded the said protection as reactive [1] if it was taken in response to a data breach notification; or [0] if the protection was not taken at all.

Protections	Measurement Items and Response Scale	Sources
	Instruction: Now I am going to ask you about any actions taken to prevent someone from obtaining your personal information. Please consider whether you have taken any of these actions during the past 12 months (1=Yes, 0=No):	
Coding of proactive and reactive protections		
	The measurement of each protection has two parts as shown below. Part (a) asks whether a respondent took the protection or not; and if the respondent took the protection, part (b) asks whether the protection was taken reactively in response to any previous misuse of PII (e.g., a data breach notification informing the respondent about PII leakage in a data breach). If a protection was adopted before any misuse of PII, it is coded as proactive [1]; or [0] otherwise. If a protection was adopted in response to any previous misuse of PII, it is coded as reactive [1]; or [0] otherwise.	Burnes et al 2020.
Preventive protections		
SecurePC	a) Have you used any type of security software program on your computer to protect it against loss of credit cards/card theft? b) Did you do this in response to any previous misuse of your personal information?	Burnes et al 2020; Reyns and Henson 2016; Lai et al 2012.
Change Password	a) During the past 12 months, have you changed passwords on any of your financial accounts? b) Did you do this in response to any previous misuse of your personal information?	Burnes et al 2020; Reyns and Henson 2016; Milne 2003.
ShredDoc	a) During the past 12 months, have you shredded or destroyed documents that contained personal identifying information? b) Did you do this in response to any previous misuse of your personal information?	Burnes et al 2020; Lai et al 2012; Milne 2003.
Detective protections		
PurchaseIDT protection	a) During the past 12 months, have you purchased identity theft protection from a company that offers protection services? b) Did you do this in response to any previous misuse of your personal information?	Burnes et al 2020; Lai et al 2012.
Check Statement	a) During the past 12 months, have you checked your banking or credit card statements for unfamiliar charges? b) Did you do this in response to any previous misuse of your personal information?	Burnes et al 2020; Lai et al 2012; Milne 2003.
Check CreditReport	a) During the past 12 months, have you checked your credit report? b) Did you do this in response to any previous misuse of your personal information?	Burnes et al 2020; Lai et al 2012; Milne 2003.

Table 17: Measurement instruments, sources, and summary statistics of independent variables.

9.2.3 Control Variables

We control for the antecedents of IDT victimization identified by previous IDT studies summarized in Tables 14 and 15. Specifically, we control for individual demographics, education level, employment status, household income (Anderson 2006;

Burnes et al. 2020; Reyns and Henson 2016), online shopping behavior, frequency of online purchase (Burnes et al. 2020; Goel 2019; Reyns and Henson 2016); property victimization, violent victimization, or identity theft victimization in the previous year (Burnes et al. 2020; Clay-Warner et al. 2016); and leakage of social security number (SSN) in an organizational data breach (*OrgPIILeak_SSN*) (Burnes et al. 2020). Data for these controls came from ITS in year *t* and NCVS in year (*t – 1*). Table 18 presents the measurement items of the controls.

Individual Characteristics	Measurement Item and Response Scale	Sources
Online Behavioral Characteristics		
Online purchase	During the past 12 months, have you used the Internet to purchase anything online? (1 = Yes, 0 = No)	Burnes et al 2020; Reyns and Henson 2016;
Frequency of online purchase	About how many times did you purchase something online, during the past year?	
Socio-demographics		
Age	Age at last birthday	Clay-Warner et al 2016; Bunch et al 2014; Copes et al 2010; Anderson 2006;
Male	1 if gender is male, 0 otherwise	
White	1 if race is White, 0 otherwise	
Black	1 if race is Black, 0 otherwise	
College degree	1 if education attainment is equal to or beyond a college degree, 0 otherwise	
Employed	1 if respondent has a job during the last 6 months, 0 otherwise	
Household income	Total combined income of all household members in any of the following 14 categories: 1) Less than $5,000; 2) $5,000 to $7,499; 3) $7,500 to $9,999; 4) $10,000 to $12,499; 5) $12,500 to $14,999; 6) $15,000 to $17,499; 7) $17,500 to $19,999; 8) $20,000 to $24,999; 9) $25,000 to $29,999; 10) $30,000 to $34,999; 11) $35,000 to $39,999; 12) $40,000 to $49,999; 13) $50,000 to $74,999; 14) $75,000 or more	
Property victimization	1 if something belongs to the respondent was stolen during the past 12 months, 0 otherwise	Burnes et al 2020; Clay-Warner et al 2016; Bunch et al 2014
Violent victimization	1 if respondent was attacked or threatened during the past 12 months, 0 otherwise	
Prior IDT victimization	1 if a respondent experienced IDT outside of the past 12 months, 0 otherwise	

Table 18: Measurement instruments, sources, and summary statistics of control variables.

9.3 Research Design and Sample Construction

9.3.1 Overview

The observational data in NCVS and ITS were not generated by experiments. An inherent problem of data analysis using non-experimental data is the potential selection

bias. In our research context, it means that individuals may self-select into taking the protections based on their pre-existing risk profiles. Therefore, the observed effectiveness of individual protections may be contaminated by the risk profiles. For instance, wealthier individuals who have more money may be more likely to take the protections because they may be higher value targets for criminals. Indeed, there is evidence that individuals who perceive higher PII leakage risk or IDT victimization risk are more likely to adopt the protections (Lai et al. 2012). If a study does not account for such potential self-selection bias, its results could also be biased. None of the empirical IDT studies we reviewed accounted for potential self-selection bias.

Our research design includes two approaches to address the potential selection bias. First, the longitudinal structure of the NCVS-ITS data allows us to track the same respondent for two ITS periods. With repeated observation of the same respondent, we use the fixed-effects regression to control for unobserved time-invariant individual characteristics which may confound with the effectiveness of protections. Second, the NCVS contains a rich set of individual characteristics. We use this information to create a matched cross-sectional sample using the propensity score matching (PSM) method. The central idea of PSM is to approximate an experiment by pruning the sample in the treatment and control groups. Respondents in the treatment group take protections while those in the control group do not. The matched sample retains respondents in both groups with the same estimated probability of taking protections. This approach reduces the selection bias by improving the individual risk profiles based on observed characteristics.

The two approaches of fixed-effects model and PSM are complementary to each other in three aspects. First, the fixed-effects model controls for both observable individual characteristics and unobservable time-invariant individual characteristics. In comparison, the PSM only controls for observable individual characteristics. The advantage of fixed-

effects model is that it requires a less restrictive assumption for making inference. Second, the fixed-effects model uses within subject variation across time while data analysis based on PSM uses between subject variation. The loss of information is reduced by using both approaches. Third, the NCVS-ITS panel dataset only contains two observations per respondent. However, at least the third time period is required to examine whether changes of detective protections across the first two time periods would mitigate IDT risk of data breach in the third time period. Nevertheless, we examine the effectiveness of detection protections using between subject variation in the PSM sample. Therefore, data analysis with both approaches can increase the credibility of the results.

9.3.2 Panel Data Sample Construction

The panel data sample consists of repeated observations of each respondent to ITS across time periods. However, the original ITS dataset tracks households instead of individual respondents due to the survey administration. The ITS data collection is based on a rotation panel design. In this design, a national representative sample of households are selected for interview in a three-year period. After three years, the existing households in the panel groups are rotated out and a new sample of households are selected for interview. Since the ITS is conducted every two years, each selected household is interviewed for two ITS periods.

The ITS data provide detailed information about the interview process; this information contains four criteria of tracking respondents across time. First, the panel groups of households should be consistent across time. When the panel group changes, previous households are rotated out and a different sample of households will be selected. Second, residents of a household should not move across time periods because ITS interviewers only keep one address of each household. Third, household membership

should remain the same across time periods. Each respondent is assigned to a household member number, which will become inconsistent if the membership changes across time periods. Examples of membership change are someone leaves or joins the household due to marriage or college education. In addition, a respondent should be available to take the interviews by himself/herself. When a respondent is unavailable, another household member takes the interview as a proxy. Response provided by a proxy can be unreliable. We construct individual IDs for the respondents who satisfy the four criteria of consistent panel group, address, household membership, and self-response.

Sample-0: Panel Data Sample Table 19 describes the four-step sampling construction process of the panel dataset. The sampling criteria and change of sample sizes are explained at each step. In addition, we further adjust the sample size in step 5 by excluding observations with missing data. The sample size for the final panel dataset used in regression analysis is 35,169.

Data extraction criteria	Remaining sample size
Original 2012, 2014 and 2016 NCVS-ITS data	296,575
Drop observations with non-trackability due to non-availability for interview in two waves of NCVS-ITS (2012 and 2014 combination or 2014 and 2016 combination)	84,316
Drop observations that become non-trackable due to moving out of the original household address tracked by NCVS-ITS	83,544
Drop observations who become non-trackable due to changes in household composition (e.g. someone leaves marriage, etc.)	83,290
Extract observations who completed the NCVS-ITS interviews by himself/herself. Drop observations whose NCVS-ITS survey responses are provided by proxy in order to increase response credibility	72,276
Listwise delete observations that has missing values in study variables	35,169

Note: Individuals greater than 16 years are elegible for NCVS-ITS survey interview

Table 19: Panel data sample construction procedures.

9.3.3 Propensity Score Matching (PSM) Sample Construction

The PSM aims to account for the selection bias by constructing pairs of individuals who have the same propensity of taking a given protection based on their pre-existing risk profiles, but one of the individuals in the pair takes the protection while the other one does not (Rosenbaum and Rubin 1984). Individuals who take the protection form the "treatment" group. Individuals who do not take the protection form the "control" group. The use of a sample of propensity score matched pairs of individuals from the treatment and the control group aims to mimic a randomized experiment.

We use a total of 19 covariates from the NCVS in year ($t - 1$) to measure the pre-existing risk profiles of individuals (Bunch et al. 2014; Clay-Warner et al. 2016). We use

these covariates to estimate an individual's propensity of taking protections in year (t). Table 20 lists the matching covariates and their measurement items.

Individual Characteristics	Measurement Item and Response Scale	Sources
Age	Age at last birthday	Clay-Warner et al 2016; Bunch et al 2014; Copes et al 2010; Anderson 2006;
Male	1 if gender is male, 0 otherwise	
White	1 if race is White, 0 otherwise	
Black	1 if race is Black, 0 otherwise	
Hispanic	1 if respondent is from Hispanic Origin, 0 otherwise	
Married	1 if marital status is married, 0 otherwise	
College degree	1 if education attainment is equal to or beyond a college degree, 0 otherwise	
Employed	1 if respondent has a job during the last 6 months, 0 otherwise	
Non_relative	1 if one or more household member's relationship to the household head is non-relative, 0 otherwise	
Home owner	1 if respondent owns the home, 0 otherwise	
Rural	1 if the land use at the residence address is rural, 0 otherwise	
Household size	Total number of household members	
Security device	1 if respondents resident building or community has access securing devices, 0 otherwise	
Residence length	Number of years that respondent has lived at the current address	
Female headed	1 if the head of household is a female, 0 otherwise	
Household income	Total combined income of all household members in any of the following 14 categories: 1) Less than $5,000; 2) $5,000 to $7,499; 3) $7,500 to $9,999; 4) $10,000 to $12,499; 5) $12,500 to $14,999; 6) $15,000 to $17,499; 7) $17,500 to $19,999; 8) $20,000 to $24,999; 9) $25,000 to $29,999; 10) $30,000 to $34,999; 11) $35,000 to $39,999; 12) $40,000 to $49,999; 13) $50,000 to $74,999; 14) $75,000 or more	
Property victimization	1 if something belongs to the respondent was stolen during the past 12 months, 0 otherwise	Burnes et al 2020; Clay-Warner et al 2016; Bunch et al 2014
Violent victimization	1 if respondent was attacked or threatened during the past 12 months, 0 otherwise	
Prior IDT victimization	1 if a respondent experienced IDT outside of the past 12 months, 0 otherwise	

Table 20: Measurement instruments and sources of matching covariates

We estimate the propensity score conditional on the set of 19 matching covariates related to the risk profile of an individual, which may influence the individual's adoption of protections. The logistic regression used for propensity score estimation is specified as follows:

$$p_i = P(T_i = 1|X_i) = \frac{\exp[\alpha_i + \beta_i X_i + \varepsilon_i]}{1 + \exp[\alpha_i + \beta_i X_i + \varepsilon_i]} \qquad (1)$$

where p_i is the propensity score. T_i is the treatment status of individual *i*; $T_i = 1$ if individual *i* takes any protection measured by ITS in year (*t*); X_i represents the vector of 19 matching covariates of i. X_i are measured by NCVS in year (*t – 1*). The one-year lag ensures that the matching covariates are not affected by the outcome variables, because

they were fixed before individual takes any protections or experiences any PII leakage or IDT victimization event.

Sample-1: We start the sample construction process with N=296,575 observations from the pooled ITS dataset across years 2012, 2014, and 2016. Then, we pair the ITS dataset in a given year (*t*) with the corresponding NCVS dataset in the previous year (*t - 1*). The ITS respondents in year (*t*) who did not respond to NCVS in year (*t – 1*) are dropped from the sample. This yields a raw sample (Sample 1) size of N = 161,723 individuals who responded to both NVCS in (*t - 1*) and ITS in (*t*). Table 21 presents the correlations among the study variables.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 PsnPIILeak	1.00														
2 OrgPIILeak	-0.08 *	1.00													
3 OrgPIILeak_SSN	-0.04 *	0.52 *	1.00												
4 IDT Victimization	0.35 *	0.10 *	0.08 *	1.00											
5 SecurePC	0.00	0.10 *	0.06 *	0.03 *	1.00										
6 ChangePassword	-0.05 *	0.16 *	0.08 *	0.03 *	0.21 *	1.00									
7 ShredDoc	-0.10 *	0.08 *	0.03 *	0.00	0.15 *	0.24 *	1.00								
8 PurchaseIDTprotection	-0.01 †	0.05 *	0.04 *	0.01	0.18 *	0.09 *	0.08 *	1.00							
9 CheckStatements	-0.23 *	0.06 *	0.02 *	-0.05 *	0.14 *	0.22 *	0.46 *	0.07 *	1.00						
10 CheckCreditReport	-0.06 *	0.08 *	0.04 *	0.01	0.16 *	0.30 *	0.23 *	0.12 *	0.23 *	1.00					
11 Age	0.00	-0.01 *	0.00	0.00	-0.02 *	-0.09 *	0.10 *	0.02 *	0.12 *	-0.04 *	1.00				
12 Male	-0.01 *	-0.02 *	0.01	0.00	0.01 †	0.03 *	-0.03 *	0.01 *	-0.02 *	0.02 *	-0.03 *	1.00			
13 White	0.01	0.04	0.02	0.03	0.04	0.07	0.08	0.01	0.09	0.03	0.09	0.02	1.00		
14 Black	-0.01	-0.04 *	-0.01 *	-0.03 *	-0.04 *	-0.08 *	-0.07 *	0.00	-0.08 *	-0.03 *	-0.04 *	-0.03 *	-0.75 *	1.00	
15 Hispanic	-0.01 *	-0.06 *	-0.02 *	-0.04 *	-0.06 *	-0.08 *	-0.11 *	-0.03 *	-0.14 *	-0.05 *	-0.16 *	0.00	0.12 *	-0.10 *	1.00
16 Married	0.02 *	0.08 *	0.04 *	0.04 *	0.08 *	0.09 *	0.13 *	0.04 *	0.10 *	0.10 *	0.15 *	0.06 *	0.09 *	-0.12 *	-0.02 *
17 College Degree	0.00	0.00	0.01	0.01 *	0.00	0.02 *	0.01 *	0.01 *	0.01 *	0.00	0.00	0.00	-0.01 *	0.01	0.00
18 Household Size	-0.01	0.01 †	0.00	-0.01 †	0.00	0.00	-0.09 *	-0.02 *	-0.11 *	-0.01 †	-0.39 *	0.01 *	-0.05 *	-0.02 *	0.20 *
19 Non_relative	-0.01 †	-0.01	-0.01 *	0.00	-0.01	0.01 *	-0.01 *	-0.01 *	-0.01 †	0.01 *	-0.17 *	0.02 *	0.01	-0.01 *	0.01 †
20 Home Owner	0.00	0.00	0.00	0.00	0.00	-0.01	-0.01	0.01	0.00	0.00	0.00	-0.01 *	-0.01	0.01	-0.01
21 Rural	0.00	0.00	-0.02 *	-0.01	0.00	-0.02 *	-0.01 *	-0.01	0.00	-0.01 *	-0.01	0.00	0.00	0.00	-0.01
22 Employed	0.00	0.01 †	0.01 †	0.00	0.01 †	0.01	0.00	0.01	0.01 †	0.00	0.00	0.00	0.01	0.00	0.00
23 Security Device	0.01	0.00	0.01	0.01 *	0.00	0.00	0.00	0.00	-0.01	0.00	-0.01	0.00	-0.01 †	0.01	0.00
24 Residence Length	-0.01 †	0.00	-0.01	0.00	-0.01	-0.01 †	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
25 Female Headed	-0.02 *	-0.01	0.00	-0.02 *	0.00	0.00	0.01	-0.01	0.00	0.00	0.00	0.01 †	0.00	0.00	0.00
26 Household Income	0.01 *	0.01	0.01 *	0.02 *	0.00	0.01	0.01 *	0.00	0.01 †	0.00	0.01	-0.01 *	-0.01	0.00	0.00
27 Property Victimization	0.00	0.00	0.00	0.00	0.00	-0.01	-0.01 †	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00
28 Violent Victimization	0.00	0.00	-0.01	0.00	-0.01 *	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
29 Prior IDT Victimization	0.23 *	0.16 *	0.10 *	0.08 *	0.04 *	0.10 *	0.02 *	0.01 †	-0.02 *	0.02 *	0.02 *	0.00	0.06 *	-0.05 *	-0.05 *

	16	17	18	19	20	21	22	23	24	25	26	27	28	29
17 College Degree	0.00	1.00												
18 Household Size	0.25 *	0.00	1.00											
19 Non_relative	-0.28 *	0.00	0.06 *	1.00										
20 Home Owner	-0.01	0.12 *	0.01 †	0.01	1.00									
21 Rural	0.00	-0.12 *	0.01	0.00	0.13 *	1.00								
22 Employed	0.00	0.16 *	0.00	0.00	-0.04 *	-0.06 *	1.00							
23 Security Device	0.00	0.06 *	0.00	0.01 †	-0.20 *	-0.11 *	-0.01	1.00						
24 Residence Length	-0.01 *	-0.09 *	-0.01	0.00	0.33 *	0.13 *	-0.27 *	-0.13 *	1.00					
25 Female Headed	0.00	-0.02 *	0.00	0.01	-0.04 *	-0.02 *	-0.08 *	0.01 †	0.01 *	1.00				
26 Household Income	0.00	0.34 *	0.00	-0.01	0.31 *	-0.07 *	0.31 *	-0.03 *	-0.05 *	-0.13 *	1.00			
27 Property Victimization	0.01	-0.02 *	0.00	-0.01 *	-0.05 *	-0.04 *	0.02 *	-0.01	-0.04 *	0.00	-0.04 *	1.00		
28 Violent Victimization	0.00	-0.01	-0.01	0.00	-0.02 *	-0.01 *	0.00	-0.01	-0.02 *	0.01	-0.02 *	0.13 *	1.00	
29 Prior IDT Victimization	0.05 *	0.01	-0.02 *	-0.01 †	0.00	-0.01 *	0.00	0.01	-0.01	-0.01	0.00	0.00	0.00	1.00

Note: Correlation significance levels: †p < 0.1; * p < 0.05 or lower (this includes p < 0.01 and p < 0.001. We did not put multiple stars due to space limitation). N = 161,723

Table 21: Correlations of study variables.

Sample-2: Propensity Score Matched Sample for the Proactive Protection Takers. Before we estimate the individuals' propensity scores of taking proactive protections, we list wise delete all observations that have missing values in the 19 matching covariates, or the dependent, independent, or the control variables of the study. The size of the retained sample before propensity score estimation is N = 33,170. Once the propensity scores are estimated, to estimate the average treatment effect (ATE), each "treated" individual, i.e., an individual who took proactive protections, needs to be matched to a "control" individual,

who had the same propensity score of taking the protections based on her risk profile, but did not take the protections. The size of the treated group is N = 29,022 whereas the size of the control group is N = 4,148 in the pre-matching sample. In such situations, prior research recommends drawing from the control group with replacement (Dehejia and Wahba 2002). This way, one individual in the control group can be matched with multiple individuals in the treatment group. We use the nearest neighbor-matching algorithm in implementing the PSM (Kumar et al. 2018). This algorithm matches each treated individual to a different control individual. For some individual, multiple observations across years exists in the pooled cross-sectional sample. Matched pairs of the same individual are dropped (N = 9). In addition, if the difference in the propensity scores of the treated individual and the control exceeds two standard deviations of the observed distribution of the differences, we drop the matched pair (Aral et al. 2009). This matching process results in a propensity score matched sample (Sample 2) with N = 66,224 observations for the proactive protection takers.

Sample-3: Propensity Score Matched Sample for the Reactive Protection Takers. To test the effect of reactive protections on IDT victimization, we need to compare respondents who take only reactive protections with respondents who do not take any protections at all. Therefore, we drop respondents who took any proactive protections. We also drop respondents who experienced an IDT victimization in the previous year to make sure that their reactive protections were in response to a PII leakage within the current 12 month period measured by ITS. The retained sample before the propensity score estimation has N = 3,819 observations. Once the propensity scores are estimated, each treated individual who took reactive protections is matched to a different control individual who had the same propensity score but did not take any reactive protections. The size of the treatment group is N = 321 whereas the size of the control group is N = 3,498 in the pre-

matching sample. Each individual in the treatment group is matched with an individual in the control group using the nearest-matching algorithm. Matched pairs of the same individual are also dropped. This matching process yielded a propensity score matched sample (Sample 3) with N = 626 observations for the reactive protection takers.

The summary statistics for Samples 1, 2 and 3 are shown in Table 22. The mean of IDT victimization is as high as 0.34 in Sample 3 whereas it is only 0.08 in Sample 2 and 0.06 in Sample 1. Recall that Sample 3 consists of individuals who experienced PII leakage in a data breach. These descriptive statistics provide preliminary evidence that PII leakage in a data breach has a strong positive association with subsequent IDT victimization.

Variables	Sample 1: ITS matched with NCVS. N = 161,723			Sample 2: PSM matched sample for proactive protection takers. N = 66,224		Sample 3: PSM matched sample for reactive protection takers. N = 626	
	Mean	Std.Dev.	% Missing values	Mean	Std.Dev.	Mean	Std.Dev.
1 PsnPIILeak	0.05	0.21	8.79%	0.06	0.23	-	-
2 OrgPIILeak	0.07	0.26	8.79%	0.09	0.28	0.10	0.30
3 OrgPIILeak_SSN	0.02	0.14	8.76%	0.03	0.16	0.04	0.18
4 IDT Victimization	0.06	0.25	13.92%	0.08	0.26	0.34	0.47
5 SecurePC	0.12	0.33	8.83%	0.15	0.35	0.11	0.32
6 ChangePassword	0.26	0.44	8.78%	0.30	0.46	0.30	0.46
7 ShredDoc	0.58	0.49	8.59%	0.71	0.45	0.36	0.48
8 PurchaseIDTprotection	0.03	0.18	8.66%	0.03	0.18	0.07	0.26
9 CheckStatements	0.64	0.48	8.55%	0.78	0.42	0.47	0.50
10 CheckCreditReport	0.32	0.47	8.62%	0.38	0.49	0.25	0.44
11 Age	48.19	18.36	0.00%	50.00	17.50	49.13	16.52
12 Male	0.48	0.50	0.00%	0.47	0.50	0.48	0.50
13 White	0.82	0.39	0.00%	0.83	0.37	0.79	0.41
14 Black	0.11	0.31	0.00%	0.10	0.30	0.11	0.31
15 Hispanic	0.15	0.35	0.13%	0.12	0.33	0.17	0.37
16 Married	0.57	0.50	0.75%	0.59	0.49	0.59	0.49
17 College Degree	0.38	0.49	2.01%	0.45	0.50	0.42	0.49
18 Household Size	2.84	1.51	0.00%	2.66	1.42	2.78	1.55
19 Non_relative	0.08	0.27	0.00%	0.07	0.26	0.06	0.24
20 Home Owner	0.74	0.44	0.00%	0.78	0.41	0.80	0.40
21 Rural	0.18	0.39	0.00%	0.20	0.40	0.23	0.42
22 Employed	0.58	0.49	12.57%	0.61	0.49	0.55	0.50
23 Security Device	0.10	0.30	0.00%	0.09	0.29	0.11	0.32
24 Residence Length	13.71	12.69	20.07%	13.64	12.68	14.56	13.39
25 Female Headed	0.49	0.50	0.00%	0.49	0.50	0.46	0.50
26 Household Income	11.44	3.32	29.40%	11.39	3.23	11.61	3.12
27 Property Victimization	0.06	0.25	50.71%	0.06	0.24	0.07	0.25
28 Violent Victimization	0.01	0.10	11.18%	0.01	0.11	0.02	0.13
29 Prior IDT Victimization	0.09	0.29	25.35%	0.10	0.30	-	-

Table 22: Summary statistics of variables for all PSM samples.

Assessment of Covariate Balance in the PSM Samples. We next assess if the covariate distributions in our PSM samples (Sample 2 and Sample 3) have balance or high degree of overlap between the treatment and the control groups. Many empirical studies report t-statistic as the measure of covariate balance in PSM samples (Clay-Warner et al. 2016; Jung et al. 2019; Kwon and Johnson 2018). However, methodologists argue that the use of t-statistic for balance assessment in PSM could be inappropriate for two reasons (Ho et al. 2007; Imbens and Rubin 2015). First, the t-statistic is sensitive to the sample size. Second, both the differences in the means and the dispersion of the two groups should be compared when assessing the overlap in the covariate distributions. The t-statistic only captures the differences in the means; it does not capture the differences in dispersion. Previous research recommends using the following measures of covariate balance to address the limitations of the t-statistic: (1) use normalized differences (Nor.Dif), which measure the differences in means and are not sensitive to sample size; and (2) use the logarithm of the ratio of standard deviations (Log.Rat of SD), which measures the differences in dispersion (Imbens and Rubin 2015). We compare these measures for each covariate and the estimated propensity score between treatment and control groups. Table 23 presents the estimates of the two measures and the sample means before and after matching for the proactive protections. Table 24 presents the same measures for the reactive protections.

	Unmatched Sample, N = 33,170				Matched Sample, N = 66,224			
	Means			Log Rat of	Means			Log Rat of
	Treated	Control	Nor. Dif	STD	Treated	Control	Nor. Dif	STD
Age	50.57	45.72	0.26	-0.18	50.06	50.53	-0.03	0.01
Male	0.46	0.50	-0.07	0.00	0.47	0.48	-0.02	0.00
Black	0.84	0.75	-0.22	-0.25	0.83	0.83	-0.01	-0.01
White	0.09	0.16	0.23	-0.17	0.10	0.10	0.01	-0.01
Hispanic	0.11	0.23	-0.33	-0.31	0.12	0.12	0.00	0.00
Married	0.61	0.45	0.34	-0.02	0.59	0.58	0.03	0.00
College Degree	0.45	0.43	0.04	0.00	0.45	0.44	0.03	0.00
Household Size	2.62	3.04	-0.27	-0.22	2.66	2.65	0.01	0.00
Non_relative	0.07	0.08	-0.01	-0.02	0.07	0.08	-0.02	-0.03
Home Owner	0.78	0.78	-0.01	0.01	0.78	0.78	-0.01	0.01
Rural	0.19	0.21	-0.03	-0.02	0.20	0.20	-0.02	-0.02
Employed	0.61	0.61	0.01	0.00	0.61	0.61	0.00	0.00
Security Device	0.09	0.09	0.00	0.00	0.09	0.09	0.02	0.02
Residence Length	13.63	13.78	-0.01	-0.01	13.64	13.59	0.00	0.01
Female Headed	0.49	0.49	0.01	0.00	0.49	0.48	0.01	0.00
Property Victimization	11.40	11.32	-0.02	-0.03	11.39	11.38	0.03	0.05
Violent Victimization	0.06	0.07	-0.05	-0.19	0.06	0.06	0.00	0.02
Household Income	0.01	0.02	0.03	-0.03	0.01	0.01	0.00	0.00
Prior IDT Victimization	0.10	0.08	0.07	0.10	0.10	0.11	-0.05	-0.06
Multivariate Measure			0.63				0.08	
Propensity Score	0.87	0.80	0.66	-0.51	0.86	0.86	0.00	0.00

Table 23: Covariate balance before and after PSM for proactive protection takers.

	Unmatched Sample, N = 3,819				Matched Sample, N = 626			
	Means			Log Rat of	Means			Log Rat of
	Treated	Control	Nor. Dif	STD	Treated	Control	Nor. Dif	STD
Age	48.83	45.07	0.20	-0.28	48.95	49.26	-0.02	-0.04
Male	0.46	0.51	-0.08	0.00	0.48	0.50	-0.04	0.00
Black	0.79	0.74	-0.20	-0.21	0.79	0.79	0.04	0.06
White	0.11	0.17	0.13	-0.08	0.11	0.10	0.01	-0.01
Hispanic	0.16	0.25	-0.22	-0.16	0.16	0.17	-0.01	-0.01
Married	0.59	0.42	0.35	0.00	0.58	0.63	-0.10	0.02
College Degree	0.43	0.43	-0.01	0.00	0.42	0.41	0.02	0.00
Household Size	2.73	3.11	-0.23	-0.13	2.74	2.82	-0.05	0.00
Non_relative	0.07	0.08	-0.05	-0.08	0.07	0.05	0.08	0.16
Home Owner	0.79	0.78	0.02	-0.01	0.79	0.80	-0.01	0.01
Rural	0.24	0.21	0.07	0.05	0.23	0.23	0.02	0.01
Employed	0.55	0.61	-0.12	0.02	0.56	0.54	0.04	0.00
Security Device	0.11	0.09	0.07	0.09	0.10	0.12	-0.07	-0.09
Residence Length	14.75	13.70	0.08	0.12	14.54	14.54	0.00	0.07
Female Headed	0.43	0.49	-0.12	-0.01	0.44	0.47	-0.05	0.00
Property Victimization	11.56	11.29	-0.03	-0.06	11.55	11.65	-0.01	-0.02
Violent Victimization	0.06	0.07	-0.01	-0.05	0.06	0.06	-0.07	-0.23
Household Income	0.02	0.02	0.08	-0.08	0.02	0.03	-0.03	0.00
Multivariate Measure			0.61				0.20	
Propensity Score	0.12	0.08	0.73	0.05	0.12	0.12	0.00	0.00

Table 24: Covariate balance before and after PSM for reactive protection takers.

From Table 23 and 24, it is clear that the PSM improved the covariate balance. None of the estimates of normalized differences in the two tables exceeds 0.1, which is comparable to what one might expect in a randomized experiment (Imbens and Rubin 2015). In addition to comparing the univariate distributions for each covariate between treatment and control groups, we also directly compared the multivariate distributions for the set of all covariates. For the two matched samples, normalized differences of multivariate distributions do not exceed 0.2, which indicates adequate balance (Rosenbaum and Rubin 1985). Tables 23 and 24 also show that the propensity scores are well matched with estimated measures of differences of mean and dispersion equal to 0. Figure 4 presents the quantile-quantile plots (Q-Q plots) of the propensity scores before and after matching.

Q-Q plots also show that propensity scores in the matched samples are well balanced (Ho et al. 2007).

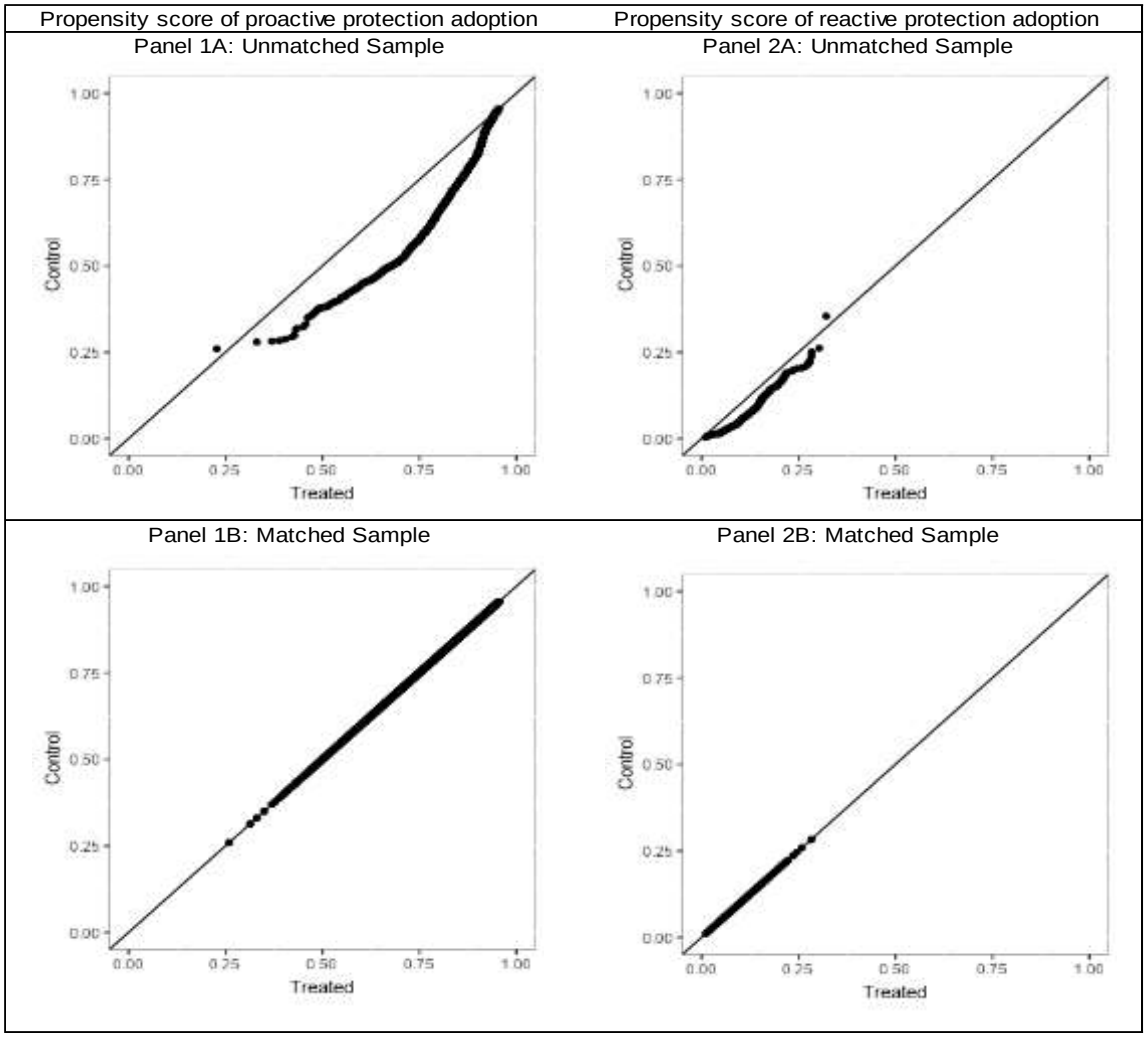

Figure 4: Q-Q plots of estimated propensity scores.

9.4 ESTIMATION MODELS AND RESULTS

9.4.1 Proactive Preventive Protections and PII Leakage

In the first stage of the IDT process, we test how proactive preventive protections affect the likelihoods of personal PII leakage and organizational PII leakage, as hypothesized in H1. The dependent variable is individual's PII leakage experience including three potential outcomes: no PII leakage, personal PII leakage, or organizational PII leakage. We model the probability of personal and organizational PII leakage with the fixed-effects logistics regression. Specifically, let $PII\ leak_{it}$ be a random variable that indicates the realized event of PII leakage of person *i* in period *t*. This variable can capture three alternative events, i.e., $PII\ leak_{it} \in \{0, 1, 2\}$; where j = 0 stands for no PII leaked, j = 1 stands for PII leaked through a personal data breach, and j = 2 stands for PII leaked through an organizational data breach. The formal statistical model is specified as follows:

$$\log\left(\frac{\Pr(PII\ leak_{it} = j)}{\Pr(PII\ leak_{it} = 0)}\right) = \alpha_{1t}^{j} + ProactivePrevs_{it}'\beta_1^j + X_{it}'\delta_1^j + \tau_1^j, \quad j = 1, 2 \qquad (2)$$

where $ProactivePrevs_{it}$ is a vector of variables for the three proactive preventive protections; X_{it} is the vector of individual time-varying socio-demographics as control variables; The terms $\{\alpha_{1t}, \boldsymbol{\beta}_1, \boldsymbol{\delta_1}\}$ are parameters to be estimated and $\boldsymbol{\tau}_1$ is the vector of time-invariant individual characteristics.

The fixed-effects logit model in equation 2 is estimated using the panel data in Sample-0. Estimates are obtained from the conditional maximum likelihood method. The results in Table 25 show that two out of the three measures of proactive preventive protections, i.e., changing passwords (Odds Ratio = 0.518, P < 0.001), and shredding personal sensitive documents (Odds Ratio = 0.332, P < 0.001), significantly reduce

personal PII leakage. Thus, H1a is partially supported. Securing personal computers (Odds Ratio = 1.141, P < 0.001) and changing passwords (Odds Ratio = 1.326, P < 0.001) significantly increase organizational PII leakage as hypothesized in H1b. However, shredding personal sensitive documents has no significant association with organizational PII leakage. Thus, H1b is supported for organizational PII leakage through digital means (e.g., computing devices), but not supported for organizational PII leakage through physical means (e.g., paper files). We also estimate the logit model with cross-sectional PSM Sample 2. Data analysis results from the PSM sample and the panel data sample are qualitatively consistent.

	PsnPIILeak vs. No PII leakage				OrgPIILeak vs. No PII leakage			
	Coef.	(Std.Dev.)		Odd ratio	Coef.	(Std.Dev.)		Odd ratio
Proactive Preventive Protections								
SecurePC	-0.058	(0.059)		0.944	0.132	(0.037)	***	1.141
ChangePassword	-0.658	(0.051)	***	0.518	0.282	(0.034)	***	1.326
ShredDoc	-1.104	(0.042)	***	0.332	0.000	(0.044)		1.000
Control variables								
Individual characteristics	Yes				Yes			
Household characteristics	Yes				Yes			
Year trend	Yes				Yes			
Note: † p < 0.1; * p < 0.05; ** p < 0.01; *** p < 0.001								

Table 25: Logistic regression estimates of PII leakage probabilities.

9.4.2 Effect of Data Breach on IDT Victimization and Moderating Role of Proactive Protections

In H2 through H4, we test whether PII leakage in a personal or an organizational data breach subsequently affects IDT victimization, and whether proactive detective protections can reduce the heightened IDT risk following the data breach. Specifically, we model the likelihood of IDT victimization using a binary logistic regression as follows:

$$\log\left(\frac{\Pr(IDT\ victimization_i)}{1-\Pr(IDT\ victimization_i)}\right)$$
$$= \alpha_2 + ProactivePrevs_i'\beta_{21} \quad (3)$$
$$+ProactiveDetcs_i'\beta_{22} + PII\ Leakage_i'\beta_{23}$$
$$+ProactiveDetcs_i \times PII\ Leakage_i{}'\beta_{24}$$
$$+X_i'\delta_2 + T_i'\tau_2 + \mu$$

where $ProactivePrevs_i$ is the vector of variables for the three proactive preventive protections; $ProactiveDetcs_i$ is the vector of variables for the three proactive detective protections; $PII\ Leakage_i$ is the vector of variables for individual *i*'s PII leakage event, including *personal PII leakage* and *organization PII leakage. *$ProactiveDetcs_i \times PII\ Leakage_i$ is a vector of interactions between each proactive detective protection and personal PII leakage and organizational PII leakage. X_i is the vector of individual socio-demographics as control variables; T_i is the vector of year dummies. The terms $\{\alpha_2, \boldsymbol{\beta}_2, \boldsymbol{\delta}_2, \boldsymbol{\tau}_2\}$ are parameters to be estimated and μ represents the error term.

The model is estimated using the cross-sectional PSM Sample-2. Ideally, a fixed-effects regression specification of this model is desirable. However, it is not applicable because the estimation of a fixed-effects model for the protection's effects on IDT victimization in the second stage requires at least observations of each individual for 3 time periods. The current NCVS-ITS survey design provides observations of each individual in a maximum of 2 time periods. Therefore, the PSM Sample-2 is the best sample available from the original data to address potential selection bias. Estimation results of equation 3 are summarized in Table 26.

Dependent Variable:	(1) Credit Card Fraud	(2) Debit Card Fraud	(3) Other Existing Account Fraud	(4) New Account Fraud	(5) Identity Theft
Proactive Preventive Protections					
SecurePC	1.037	1.016	1.18*	1.051	0.911
	(0.034)	(0.040)	(0.085)	(0.107)	(0.131)
ChangePassword	1.115***	0.912***	0.967	0.949	1.098
	(0.029)	(0.035)	(0.076)	(0.093)	(0.106)
ShredDoc	1.003	0.923**	0.968	0.712***	0.719***
	(0.031)	(0.034)	(0.076)	(0.083)	(0.101)
Proactive Detective Protections					
PurchaseIDTprotection	1.166*	0.725**	1.159	2.729***	1.416
	(0.084)	(0.128)	(0.253)	(0.243)	(0.348)
CheckStatement	2.057***	2.306***	2.594***	2.599***	2.758***
	(0.048)	(0.055)	(0.148)	(0.166)	(0.170)
CheckCreditReport	0.983	1.341***	1.448***	1.463**	1.273
	(0.039)	(0.046)	(0.117)	(0.149)	(0.156)
PII Leakage					
PsnPIILeak	21.97***	43.165***	29.376***	80.658***	36.565***
	(0.039)	(0.043)	(0.122)	(0.122)	(0.127)
OrgPIILeak	19.673***	19.7***	33.068***	65.966***	28.985***
	(0.048)	(0.057)	(0.139)	(0.142)	(0.159)
OrgPIILeak with SSN	0.996	1.461***	1.787***	1.322***	1.742***
	(0.041)	(0.049)	(0.086)	(0.096)	(0.117)
Interaction Terms					
PurchaseIDTprotection	0.644**	1.145	0.908	0.43**	0.846
× PsnPIILeak	(0.173)	(0.196)	(0.394)	(0.387)	(0.546)
PurchaseIDTprotection	0.811	1.346	0.511	0.483	0.485
× OrgPIILeak	(0.149)	(0.194)	(0.413)	(0.444)	(0.577)
CheckStatement	0.385***	0.345***	0.444***	0.367***	0.333***
× PsnPIILeak	(0.064)	(0.069)	(0.173)	(0.187)	(0.213)
CheckStatement	0.144***	0.14***	0.134***	0.082***	0.183***
× OrgPIILeak	(0.066)	(0.080)	(0.179)	(0.220)	(0.220)
CheckCreditReport	0.766***	0.757***	0.672**	0.338***	0.481***
× PsnPIILeak	(0.066)	(0.069)	(0.166)	(0.202)	(0.238)
CheckCreditReport	0.618***	0.598***	0.559***	0.295***	0.494***
× OrgPIILeak	(0.069)	(0.083)	(0.170)	(0.241)	(0.239)
Control Variables					
Individual characteristics	Y	Y	Y	Y	Y
Household characteristics	Y	Y	Y	Y	Y
Year trend	Y	Y	Y	Y	Y

Note: * p < 0.1; ** p < 0.05; *** p < 0.01;

Table 26: Odds ration estimates from logit models of proactive protections' influence on identity frauds.

The results in Table 26 show that both PII leakage through personal and organizational channel increase IDT risks. Results in Column 1 through Column 5 show that PII leakage through both channels affects all five identity frauds consistently. Specifically, individuals whose PII leak in a personal data breach are about 22 to 81 times more likely to experience varying identity frauds victimization (Odds Ratios of having 5 identity frauds are 21.97, 43.16, 29.37, 80.56, 36.56 from Column 1 to Column 5, $P < 0.01$)

than individuals who do not experience a personal data breach. Thus, H2a is supported. Further, individuals who experience PII leakage through an organizational data breach (Odds Ratios of having 5 identity frauds are 19.67, 19.70, 33.07, 65.96, 28.98 from Column 1 to Column 5, $P < 0.01$) are, on average, about 29 to 66 times more likely to experience IDT victimization compared to people who do not experience any PII leakage. These results support H2b.

Organizational PII leakage very significantly increases IDT. However, two proactive detective protections – checking bank statements and credit bureau reports - consistently mitigate the heightened risks of all identity frauds after PII leakage. The coefficient estimates of the interaction terms between these two proactive detective protections and both PII leakage channel in Column 1 to Column 5 illustrate the reduced odds ratio of all identity frauds risks. On average, proactively checking bank statements reduces the odds ratios of all heightened identity frauds risks after PII leakage from personal channel by about 65% ($P < 0.01$). It reduces even better the odds ratios of heightened identity fraud risks after PII leakage from organizational data breaches by more than 80% ($P < 0.01$). Meanwhile, proactively checking credit bureau reports reduces the odds ratios of heightened risks after PII leakage from personal channel by about 30% ($P < 0.01$) for existing account frauds and 55% ($P < 0.01$) for new account frauds and IDT. It reduces the odds ratios of heightened risks after PII leakage from organizational data breaches by a similar magnitude ($P < 0.01$). The results strongly suggest that proactively taking detective protections can reduce the heightened risks of all identity frauds after PII leakage in the second stage of the IDT process. Therefore, H4 is partially supported for proactive detective protections of checking bank statements and checking credit bureau report.

In contrast, the mitigation effects of purchasing IDT protection services on the heightened identity frauds risks are barely working. It only reduces the odds ratios of credit card fraud by 35% ($P < 0.01$) and new account fraud by 57% ($P < 0.01$) after PII leaked from personal data breach. Meanwhile, it has no mitigation effects for debit card fraud, other existing account fraud, and IDT. Opposite to the commonly promotion for commercial IDT protection services, it has no mitigation effects for all five identity frauds after PII leakage from organizational data breaches. Therefore, H4 is also partially supported for the proactive detective protection of purchasing IDT protection services.

The proactive detective protection of checking bank statement also has a strong discovery effect. The coefficient estimates of the main effect terms of checking bank statements in Table 26 from Column 1 to Column 5 are consistently positive and significant. Specifically, proactively checking bank statements increases the odds ratios of discovering all five identity frauds by 2.5 times on average (Odds ratios are 2.05, 2.31, 2.59, 2.60, 2.76 in Column 1 to Column 5, $P < 0.01$). Meanwhile, checking credit bureau reports increases the odds ratios of discovering debit card fraud (Odds ratio = 1.34, $P < 0.01$), other existing account fraud (Odds ratio = 1.45, $P < 0.01$) and new account fraud (Odds ratio = 1.46, $P < 0.01$). It does not help to discover occurrence of credit card fraud and IDT. One possible explanation is that credit card companies have their own systems of detecting frauds and individuals receive frauds notifications before they check credit bureau reports. For IDT such as job application and tax filing, the fraudulent events will not appear on credit bureau reports since these reports mainly keep records of credit histories.

On the other hand, purchasing IDT protections only help individuals discover credit card fraud (Odds ratio = 1.16, $P < 0.1$) and new account fraud (Odds ratio = 2.72, $P < 0.01$). However, it reduces the odds ratio of discovering debit card fraud by 27% ($P < 0.05$).

Compare this with the observation on the mitigation effect of purchasing IDT protection services, the results suggests that this protection serves as a detective protection only for credit card fraud and new account fraud. Since the IDT protection include a variety of services, such as protecting passwords, it may also serve as a preventive protection. The coefficient estimates on the relationship between purchasing IDT protection services and debit card fraud confirms this explanation because it reduces the odds ratio of debit card fraud. In summary, H3 is partially supported by the above results on the discovering effects of proactive detective protections.

As for proactive preventive preventions, results in Table 26 show that they have limited effects on reducing the identity fraud risks. Specifically, securing computers does not reduce the odds ratio of any identity frauds. Changing passwords of financial accounts only reduces the odds ratio of debit card fraud by about 8% ($P < 0.01$). Shredding documents containing PII is the most helpful preventive protection, which reduces the odds ratio of debit card fraud (Odds ratio = 0.923, $P < 0.05$), new account fraud (Odds ratio = 0.712, $P < 0.01$) and identity theft (Odds ratio = 0.719, $P < 0.01$). However, it has no effect on reducing the risks of credit card fraud and other existing account fraud. One explanation is that new account fraud and identity theft requires in-person verification of identity with physical documents. Therefore, shredding documents containing PII reduces the risk of leaking important identity papers in the second stage of IDT process after initial PII leakage, thus reduce the risk of new account fraud and IDT. For credit card and other existing account fraud, criminals can directly exploit the PII obtained from the initial PII leakage and commit frauds online. Therefore, digital preventive protections of securing computer and changing passwords are not helpful.

We also notice that results in Table 26 show that securing computers has a positive association with other existing account fraud, while changing passwords of financial

accounts has a positive association with credit card fraud. These results are opposite to our theoretical expectations. One possible explanation is that individuals who frequently use security software on computers and change passwords are also people who have higher income and have more to lose. However, our data do not have enough granularity for the income information of respondents. Respondents who have more than $75,000 annual household income are all measured as the same income group. However, respondents in this group can have a large variance on the actual income.

9.4.3 Reactive Protections and IDT Victimization

In H5, we test whether taking protections reactively in response to a data breach notification can reduce subsequent IDT victimization risk. We model the likelihood of IDT victimization using a binary logistic regression as follows:

$$\log\left(\frac{\Pr(IDT\ victimization_i)}{1-\Pr(IDT\ victimization_i)}\right) = \alpha_3 + ReactivePrevs_i'\beta_{31} + ReactiveDetcs_i'\beta_{32} + PII\ LeakageType_i'\beta_{33} + ReactiveDetcs_i \times PII\ LeakageType_i{}'\beta_{24} + X_i'\delta_3 + T_i'\tau_3 + \mu \quad (4)$$

where $ReactivePrevs_i$ is the vector of variables for three *reactive* preventive protections; $ReactiveDetcs_i$ is the vector of variables for three *reactive* detective protections; $PII\ LeakageType_i$ captures the type of PII leakage ([1]=organizational PII leakage; [0]=personal PII leakage)[2]; $ReactiveDetcs_i \times PII\ LeakageType_i$ is a vector of interaction variables between each reactive detective protection and type of PII leakage

[2] In the reactive protection sample, all individuals experienced a data breach. We code if the breach was an organizational data breach or a personal data breach.

respectively; X_i is the vector of individual socio-demographics as control variables; and T_i is the vector of year dummies. The terms $\{\alpha_3, \boldsymbol{\beta}_3, \boldsymbol{\delta}_3, \boldsymbol{\tau}_3\}$ are parameters to be estimated and μ represents the error term. Table 27 summarizes the estimation results.

Reactive Preventive Protections	**Coef. (Std.Dev.)**		**Odds-ratio**
SecurePC	-0.315 (0.400)		0.729
ChangePassword	1.317 (0.323)	***	3.731
ShredDoc	-0.138 (0.327)		0.871
Reactive Detective Protections			
PurchaseIDTprotection	-1.311 (0.491)	**	0.270
CheckStatement	2.983 (0.369)	***	19.756
CheckCreditReport	0.609 (0.355)	†	1.839
PII Leakage			
PIILeakType	0.978 (1.537)		2.658
OrgPIILeak with SSN	0.550 (0.805)		1.733
Interaction Terms			
PIILeakType × PurchaseIDTprotection	0.877 (1.239)		2.404
PIILeakType × CheckStatement	-0.964 (1.449)		0.381
PIILeakType × CheckCreditReport	-0.573 (0.789)		0.564
Control Variables			
Individual characteristics	Yes		
Household characteristics	Yes		
Year trend	Yes		
Constant	-4.217 (0.843)	***	0.015
Model Statistics			
N. Obs.			626
Log Likelihood			-215.00
McFadden's Pseudo R-squared			0.456
Note: † p < 0.1; * p < 0.05; ** p < 0.01; *** p < 0.001			

Table 27: Estimates from logit model of reactive protections influence on overall IDT.

The results on the interaction terms in Table 27 indicate that the effectiveness of reactive protections is not significantly different in organizational versus personal PII leakage. The main effects of the reactive protections on IDT indicate that two of the three reactive preventive protections, i.e. securing personal computers and shredding personal sensitive documents, do not have any significant effects on IDT victimization. The third one, changing passwords (Odds Ratio = 3.731, P < 0.001), significantly increases IDT victimization rather than decreasing it.

As for the reactive detective protections, purchasing IDT monitoring and protection services (Odds Ratio = 0.270, P < 0.01) has a negative direct effect on IDT victimization, suggesting a *mitigation effect*. The other two protections, i.e., checking bank and credit card statements (Odds Ratio = 19.756, P < 0.001) and checking credit report (Odds Ratio = 1.839, P < 0.1), have positive and significant direct effects on IDT victimization, suggesting *discovery effects*.

9.5 Robustness Analyses

9.5.1 Sensitivity Analysis for Unobservable Variables

The propensity score matching method we use to address the selection problem relies on the assumption that observable characteristics, such as the 19 covariates we used, fully account for selection of individuals into treatment and control groups. Although the 19 covariates capture multiple aspects of an individual's risk profile, an individual may have other, unobserved characteristics that could also potentially affect her propensity to take protections. To test whether our empirical results are sensitive to such unobservable variables, we conduct sensitivity analysis by estimating Rosenbaum bounds (Leuven and Sianesi 2003; Rosenbaum 2002; Sun and Zhu 2013). Rosenbaum bounds measure how strongly the unobservable variables must influence the selection process to undermine the causal effects identified in the propensity score matching analysis. We find that an unobservable variable would have to change the odds of selection into taking *proactive* protections by at least 32% for the effects of the *proactive* protections to disappear for the outcome variable. This threshold is above the critical values (10% to 15%) reported in prior research (DiPrete and Gangl 2004). It is also on the same order of magnitude (50%) reported in other studies (Sun and Zhu 2013). For the *reactive* protections, we find that an unobservable variable would have to change the odds of selection into taking *reactive*

protections by more than 30 times for the effects of *reactive* protections to disappear for the outcome variable. These sensitivity analyses indicate that the results obtained from the propensity score matching approach are not likely to be sensitive to unobservable variables.

9.5.2 Sensitivity Analysis for Outliers, Separation, and Scale

The estimates of organizational PII leakage and personal PII leakage on the risk of IDT victimization in Table 26 are noticeably large: 29 times and 36 times respectively. While we hypothesized such effects, we also check if the large estimates could be artifacts of any statistical issues. Specifically, maximum likelihood fitting of logistic regressions could potentially be sensitive to influential outliers, separation, and inappropriate scale in the data. We examine each of these possibilities next.

Outliers can be influential to MLE and result in bad fitting. We examine in two steps whether the large odds ratios result from influential outliers. In step one, we examine the empirical distributions of all model variables. We do not find observations with extreme values. In step two, we compute the residues of predicted observations and exclude 166 out of 66,224 observations that have residue values three standard deviations away from the mean (Pregibon 1981). We then re-estimate the regression models and obtain similar results. These findings rule out the possibility that large odd rations are due to outliers.

Separation occurs when the outcome variable can be solely predicted by one predictor variable (Heinze and Schemper 2002). We tabulate each predictor variable with the outcome variable and do not find any separation issues. Thus, separation issue is unlikely to drive the large odds ratios.

Scale of variables, when measured inappropriately, can lead to very large estimates. Our model variables are all categorical except for age and frequency of online shopping. We rescale age and online shopping frequency by using the logarithm and re-estimate the

models. The results remain qualitatively the same. Thus, scale of variables is unlikely to drive the large odds ratios.

In summary, the sensitivity analyses suggest that the large odds ratios are evidence of strong associations between data breaches and IDT victimization as hypothesized.

Chapter 10: Discussions and Conclusion

The theory and empirical findings of this study advance our understanding of: (1) how PII leakage through personal or organizational data breaches subsequently affect IDT victimization risks of the data breach victims; and (2) which protections of individuals are effective in reducing the PII leakage and the IDT victimization risks. Table 28 summarizes the key findings of the study. We discuss the implications of these findings for research, practice, and future work.

Hypotheses	Dependent Variables	Independent Variables	Estimated Effects	Conclusion
Proactive Protections				
H1a	PsnPIILeak	SecurePC ChangePassword ShredDoc	*ns* - -	Partially supported.
H1b	OrgPIILeak	SecurePC ChangePassword ShredDoc	+ + *ns*	Partially supported.
H2a	IDT Victimization	PsnPIILeak	+	Supported.
H2b	IDT Victimization	OrgPIILeak	+	Supported.
H3	IDT Victimization	PurchaseIDTprotection CheckStatement CheckCreditReport	*mixed* + +	Partially supported.
H4a	IDT Victimization	PsnPIILeak × PurchaseIDTprotection PsnPIILeak × CheckStatement PsnPIILeak × CheckCreditReport	- - -	Partially supported.
H4b	IDT Victimization	OrgPIILeak × PurchaseIDTprotection OrgPIILeak × CheckStatement OrgPIILeak × CheckCreditReport	*ns* - -	Partially supported.
Reactive Protections				
H5a	IDT Victimization	SecurePC ChangePassword ShredDoc	*ns* + *ns*	Not supported.
H5b	IDT Victimization	PurchaseIDTprotection CheckStatement CheckCreditReport	- + +	Supported.
Note: (*ns)* : not significant; (-): negative and significant; (+): positive and significant.				

Table 28: Summary of hypotheses testing results.

10.1 CONTRIBUTIONS TO RESEARCH

Data breach and IDT. Our starting premise that the leakage of PII in a personal or an organizational data breach significantly increases IDT victimization has received strong support. If a person's PII leaks in a personal data breach, the IDT risk of the person increases by more than 36 times, on average. If a person's PII leaks through an

organizational data breach, the IDT risk of the person increases by about 28 times, on average.

Prior academic studies advanced mixed arguments on the link between data breach and IDT. Some studies argued that PII leakage in a data breach would increase IDT risk (Roberds and Schreft 2009; Romanosky et al. 2011). Others argued that the leaked PII would not necessarily be misused subsequently to commit the IDT crime (Acquisti et al. 2016; Mann 2015). The industry practitioners also advance mixed arguments about the link between data breach and IDT. Organizational data breach notification letters uniformly warn data breach victims that they face heightened risks of IDT victimization because their PII leaked in the data breach. However, the U.S. courts do not accept the asserted link between the data breaches and IDT victimizations (Lorio 2017; Mank 2017). When the data breach victims sue the breached organizations for IDT-related injuries, the U.S. courts often deny them standing by arguing that: (i) there is no traceable causation between organizational data breaches and IDT victimization; (ii) there is no concrete injury to data breach victims; and (iii) plaintiffs' alleged injuries from organizational data breaches are speculative, hypothetical, and too reliant on subjective fears and anxieties (Solove and Citron 2018). Our theory and findings provide strong support for the link between data breaches and IDT victimization. In a nationally representative, propensity score matched sample of 66,224 citizens in the U.S.A, we find that the IDT victimization risk of a citizen whose PII leaks in an organizational data breach increases by more than 28 times, on average, compared to a citizen whose PII does not leak in a data breach. Each IDT victim in our sample suffered a concrete injury such as financial fraud, theft, or some other criminal activity that harmed the victim. Our findings imply that both the concrete injury and the traceable causation criteria of the Article III's standing criteria are likely to be met for the average victim of organizational data breaches. To our knowledge, this is the first

academic study that provides large sample empirical evidence on the contentious link between data breaches and IDT victimizations.

Protections and IDT Victimization. Prior studies viewed IDT victimization as a single-stage event and focused on how a given protection is associated with IDT victimization. Our conceptualization of IDT as a multi-stage process enables us to theoretically distinguish the roles of proactive versus reactive protections on the one hand, and preventive versus detective protections on the other. These distinctions generate new theoretical insights and explanations as to which protections might be effective in which stages of the overall IDT process.

Effectiveness of proactive versus reactive protections. As summarized in Table 28, we find strong support that proactive protections are effective in reducing PII leakage as well as IDT victimization (H1-H4). In comparison, reactive protections have mixed effects on IDT victimization. Reactive detective protections such as checking credit card and bank statements help discover the IDT victimization (H5b). Reactively taking the IDT monitoring and protection services also help with the mitigation of IDT risks. However, reactive preventive protections such as securing computers and shredding documents do not have any significant effects on IDT victimization (H5a).

Changing passwords in reaction to a data breach notification. A prior study found that changing passwords frequently reduced the likelihood of IDT victimization (Burnes et al. 2020). Our results do not replicate this finding. On the contrary, we find that changing account passwords in reaction to a data breach notification significantly increases the risk of IDT victimization. To remind, we distinguish proactive versus reactive protections and their roles in the different stages of the IDT process. We find that changing account passwords proactively is effective in reducing PII leakage in personal data breach (H1a), but it does not have any significant effect on IDT victimization. Changing account

passwords reactively in response to a PII leakage in an organizational data breach increases the victim's likelihood of subsequently experiencing IDT victimization (H5a). We believe this is because many victims tend to reuse old passwords, and password reuse enables criminals to take over additional accounts of the victim.

After compromising a person's account in an organizational data breach, criminals also attempt to take over additional accounts of the person by conducting "credential stuffing attacks." Criminals know that many people reuse their old passwords when changing passwords in current accounts in response to a data breach notification. Thus, the criminals use automated bots to match the credentials of the compromised account to previously breached accounts and cracked passwords that are sold on the dark web. If a victim reuses one of those previously cracked passwords in changing the passwords of the current accounts, criminals are able to take over the current accounts as well. Criminals can also install key loggers and remain stealth in compromised devices to capture newly changed passwords and breach additional accounts. Breaching multiple accounts of the same victim increases the IDT victimization risk by providing the criminals with sufficient PII and sufficient accounts where they can commit illegal transactions to complete the IDT victimization process. Thus, our findings suggest that the prevailing assumption about the link between changing passwords and IDT victimization ought to be revised as follows. If a victim changes passwords reactively, after receiving a data breach notification letter, changing passwords increases the person's IDT victimization risk. If the person proactively changes passwords before any PII leaks in a data breach, it reduces the IDT victimization risk of the person.

Effectiveness of preventive versus detective protections. The findings support our thesis that preventive and detective protections are effective in different stages of the IDT process. Preventive protections are effective in the earlier stages of the IDT process: they

reduce the risks of PII leakage in data breaches. Detective protections are effective in the later stages of the IDT process. Once the PII leaks in a data breach, detective protections are effective in discovering if the PII leaked, if it is being misused, alerting the victim about it, and enabling the victim to stop the criminals from proceeding further in the IDT process. In case all these efforts fail and IDT victimization occurs, the detective protections are effective in discovering that IDT victimization occurred and alerting the victim about it so that the victim can take action to contain the damages caused by the IDT victimization.

Prior studies found that preventive protections such as using antivirus software, deleting emails, and changing passwords were not effective in reducing IDT victimization (Reyns and Henson 2016). We replicate this result in our IDT regressions focusing on the last stage of the IDT process: fraud, theft, or other crime. However, we also find that these preventive protections are effective in an earlier stage of the IDT process: they significantly reduce the PII leakage through personal data breaches. Thus, our findings revise the existing assumptions about the effectiveness of preventive protections as follows: preventive protections are effective in reducing PII leakage through personal breaches but they are not effective in reducing fraud, theft, or other crimes.

As hypothesized, proactive preventive protections also affect the likelihood of organizational data breaches significantly. Individuals who secure their personal computing devices and change their financial account passwords frequently are at higher risk of experiencing organizational data breaches. This finding supports our explanation that criminals are likely to try to access the PII through organizational systems when the individuals secure the PII on their personal devices. The only preventive protection of the individuals that reduces the organizational data breach risk is the shredding of sensitive personal documents. Improperly disposed sensitive documents of a person could provide the criminals with the much needed personal details to breach the person's organizational

accounts. We find that individuals are able to disrupt the criminals' efforts to breach organizational accounts by shredding their personal documents containing PII.

Prior studies also found that IDT monitoring and protection services, one of the most commonly recommended protections against IDT risk, increased IDT victimization risk rather than decreasing it (Burnes et al. 2020). Our study replicates this finding by showing that the direct effect of the IDT monitoring and protection services on IDT victimization is positive and significant. However, our theory goes further to clarify that this is only the *discovery effect* of the IDT monitoring and protection services. Victims who use these services are able to discover and report the IDT victimization cases more than victims who do not use these services. In addition, our results show that the IDT monitoring and protection services also have a *mitigation effect*: they reduce the heightened IDT risks of data breaches. Thus, our theory and results suggest revising the existing assumptions as follows: IDT monitoring and protection services mitigate the heightened IDT risks of data breaches; and when they cannot mitigate the IDT risk, IDT monitoring and protection services enable victims to discover that IDT victimization occurred.

Methodologically, this is the first IDT study that uses a large, nationally representative sample of U.S. citizens to create a propensity score matched sample of treatment-control pairs to strengthen its causal inferences about the linkages among protections, data breaches, and IDT.

In summary, the empirical results validate the theory proposed in Figure 1b. This theory can serve as an individual-level protection framework for IDT victimization risk. Unlike prior studies, which view IDT as a single-stage event focusing only on fraud, the proposed theory conceptualizes IDT as a multi-stage process starting with PII leakage in a data breach, advancing with PII misuse, and culminating in an illegal activity such as fraud, theft, or other crime. The theory explains which individual protections are effective in

disrupting the malicious activities of the criminals in which stages of the overall IDT process, and in ultimately preventing the criminals from successfully committing the IDT crime.

10.2 CONTRIBUTIONS TO PRACTICE

Our finding of a strong and highly robust link between data breaches and IDT victimization has implications for individuals, organizations, and U.S. courts.

If a person's PII leaks through a personal data breach, the IDT risk of the person increases by more than 36 times, on average. To remind, this finding is from years 2012-2016. These risks are estimated to be much higher at this time, under the disruptions and challenges of an ongoing pandemic. Due to the sudden, forced shift of many individuals to working from home practices, criminals can find more opportunities to breach personal computing devices of the individuals and subsequently commit the IDT crime. Masses of individuals are trying to stay at home and rely on digital technologies to work from home, shop from home, learn from home, receive healthcare services online etc. They are at higher risk of personal data breaches and IDT.

On the mitigation side, our findings imply that proactive protections can enable individuals to mitigate data breach and IDT risks. Taking proactive preventive protections significantly reduces the PII leakage through personal data breaches. Interestingly, when individuals take preventive protections to protect the PII residing in their personal computing devices, the likelihood of PII leakage through organizational data breaches goes up. This finding confirms that criminals prefer to gain access to the PII by going through the weakest link or the path of least resistance. Individuals do not have full control of the PII they are required to share with government agencies and private organizations as a pre-condition of doing business with them. However, our findings indicate that individuals can

still mitigate their IDT risks if their PII leak through an organizational data breach. Individuals who take detective protections are able to significantly reduce the IDT risks following the leakage of PII through an organizational data breach. Our findings make a strong case to individuals for taking both preventive and detective protections proactively.

Our finding that organizational data breaches significantly increase the IDT risk by more than 28 times, on average, has significant implications for organizations holding PII as well as the U.S. courts that make standing decisions on data breach litigation cases. Organizations need to recognize that their failure to protect PII creates significant IDT risks for the owners of the PII. Accordingly, organizations should assume more accountability for the protection of the PII they hold. As for the U.S courts, the finding implies that the courts should reconsider their practice of denying standing to data breach victims who want to sue breached organizations for IDT-related harms (Lorio 2017). This study shows strong empirical evidence that both the concrete injury and the traceable causation criteria of the Article III's standing criteria are likely to be met for victims of organizational data breaches who sue the breached organizations for IDT-related injuries.

Finally, our findings indicate that changing passwords frequently could prove effective or ineffective depending on when in the overall IDT process the individual performs this protection. Changing passwords proactively reduces the individual's data breach and PII leakage risks. Changing passwords reactively in response to a data breach notification increases the IDT victimization risk of the individual. Our findings suggest that individuals should change passwords proactively before any PII leakage and refrain from doing it reactively in response to PII leakage.

10.3 LIMITATIONS AND FUTURE RESEARCH

This study was able to study the protections of individuals against IDT victimization risk. It had no data on the protections of organizations with which those individuals do business. Both individual and organizational protections can affect IDT victimization risks of individuals. Future research can study how individual and organizational protections interact in affecting the IDT victimization risks of individuals.

The study used measures from an archival data set collected by the U.S government. Thus, it is subject to the general limitations of an archival data set. The measurement instruments of the government's ITS survey are consistent with the protection instruments used in prior academic studies. Nevertheless, any remaining limitations of the ITS survey will need to be addressed in future academic studies.

The ITS survey focused on a total of six protections. Those protections align very well with the most commonly recommended protections in practice and the most commonly studied protections in academic research, as summarized in Table 2. Nevertheless, there are many more protections that were not covered by the ITS survey: e.g., biometrics, machine learning and artificial intelligence based protection software, etc. Future research can test the effectiveness of additional new preventive and detective protections that were not covered by the ITS survey.

The ITS survey is able to measure only the known data breaches. If an organization is not aware that it was breached, it cannot notify the individual victims of the data breach. Accordingly, the ITS survey cannot capture such unknown organizational data breaches. This is a common limitation of all data breach studies that rely on publicly known data breaches, and our study is no exception.

Finally, although we used the largest dataset available on IDT, in propensity score matching, we noticed that, for proactive protections, the size of the treatment group was

larger than that of the control group. We addressed this limitation by following prior suggestions to draw control units with replacement (Dehejia and Wahba 2002). When additional new ITS survey data become available in the future, future studies might be able to address this limitation.

References

Abagnale, F. W. 2013. *Stealing Your Life : The Ultimate Identity Theft Prevention Plan*. New York: Broadway Books.

Acquisti, A. 2004. "Privacy and Security of Personal Information," in *Economics of Information Security*. Springer, pp. 179-186.

Acquisti, A. 2008. "Identity Management, Privacy, and Price Discrimination," *IEEE security & privacy* (6:2), pp. 46-50.

Acquisti, A., Taylor, C., and Wagman, L. 2016. "The Economics of Privacy," *Journal of Economic Literature* (54:2), pp. 442-492.

Addington, L. A. 2005. "Disentangling the Effects of Bounding and Mobility on Reports of Criminal Victimization," *Journal of quantitative criminology* (21:3), pp. 321-343.

Albrecht, C., Albrecht, C., and Tzafrir, S. 2011. "How to Protect and Minimize Consumer Risk to Identity Theft," *Journal of Financial Crime* (18:4), pp. 405-414.

Allison, P. D. 2014. *Event History and Survival Analysis / Paul D. Allison*, (Second edition. ed.). Los Angeles: SAGE.

Anderson, K. B. 2006. "Who Are the Victims of Identity Theft? The Effect of Demographics," *Journal of Public Policy & Marketing* (25:2), pp. 160-171.

Anderson, K. B., Durbin, E., and Salinger, M. A. 2008. "Identity Theft," *Journal of Economic Perspectives* (22:2), pp. 171-192.

Angrist, J. D., and Pischke, J.-S. 2008. *Mostly Harmless Econometrics : An Empiricist's Companion* Princeton, NJ: Princeton University Press.

Anteby, M. 2008. "Identity Incentives as an Engaging Form of Control: Revisiting Leniencies in an Aeronautic Plant," *Organization science (Providence, R.I.)* (19:2), pp. 202-220.

Aral, S., Muchnik, L., and Sundararajan, A. 2009. "Distinguishing Influence-Based Contagion from Homophily-Driven Diffusion in Dynamic Networks," *Proceedings of the National Academy of Sciences* (106:51), pp. 21544-21549.

Ashforth, B. E. 2001. *Role Transitions in Organizational Life : An Identity-Based Perspective*. Mahwah, N.J: Lawrence Erlbaum Associates.

Bachman, R., and Taylor, B. H. 1994. "The Measurement of Family Violence and Rape by the Redesigned National Crime Victimization Survey," *Justice quarterly* (11:3), pp. 499-512.

Bandalos, D. L. 2018. *Measurement Theory and Applications for the Social Sciences / Deborah L. Bandalos*. New York, New York ;: The Guilford Press.

Barron, J. M., and Staten, M. 2003. "The Value of Comprehensive Credit Reports: Lessons from the Us Experience," *Credit Reporting Systems and the International Economy* (8), pp. 273-310.

Becker, G. 1968. "Crime and Punishment: An Economic Approach," *The Journal of Political Economy* (169), pp. 176-177.

Bennett, R. R. 1991. "Routine Activities: A Cross-National Assessment of a Criminological Perspective," *Social Forces* (70:1), pp. 147-163.
Berghel, H. 2000. "Identity Theft, Social Security Numbers, and the Web," *Communications of the ACM* (43:2), pp. 17-21.
Blumstein, A., Cohen, J., and Nagin, D. 1978. *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates*. Washington: National Academy of Sciences.
Bonneau, J., Preibusch, S., and Anderson, R. "A Birthday Present Every Eleven Wallets? The Security of Customer-Chosen Banking Pins." Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 25-40.
Bose, I., and Leung, A. C. M. 2013. "The Impact of Adoption of Identity Theft Countermeasures on Firm Value," *Decision Support Systems* (55:3), pp. 753-763.
Bose, I., and Leung, A. C. M. 2019. "Adoption of Identity Theft Countermeasures and Its Short-and Long-Term Impact on Firm Value," *MIS Quarterly* (43:1), pp. 313-327.
Bunch, J., Clay-Warner, J., and McMahon-Howard, J. 2014. "The Effects of Victimization on Routine Activities," *Criminal Justice and Behavior* (41:5), pp. 574-592.
Burke, P. J. 1991. "Identity Processes and Social Stress," *American sociological review* (56:6), pp. 836-849.
Burnes, D., DeLiema, M., and Langton, L. 2020. "Risk and Protective Factors of Identity Theft Victimization in the United States," *Preventive Medicine Reports* (17:101058).
CalderSecurity. a. "What Burglars Steal and Where They Look." Retrieved June 11th, 2021, from https://www.caldersecurity.co.uk/what-burglars-steal-and-where-they-look/#Personal_information
CalderSecurity. b. "Burglary Facts - Key Facts and Statistics on Burglaries Thoughout the Uk." Retrieved June 11th, 2021, from https://www.caldersecurity.co.uk/burglary-facts-uk/
Camp, J. L. 2004. "Digital Identity," *IEEE Technology and Society Magazine* (23:3), pp. 34-41.
Cantor, D., and Lynch, J. P. 2005. "Exploring the Effects of Changes in Design on the Analytical Uses of the Ncvs Data," *Journal of quantitative criminology* (21:3), pp. 293-319.
Cheney, J. S. 2005. "Identity Theft: Do Definitions Still Matter?," *FRB of Philadelphia Payment Cards Center Discussion Paper*).
Cheney, J. S., Hunt, R. M., Mikhed, V., Ritter, D., and Vogan, M. 2014. "Consumer Use of Fraud Alerts and Credit Freezes: An Empirical Analysis." Federal Reserve Bank of Philadelphia, Payment Cards Center Discussion Paper 14-04.
Chevalier, S. 2021. "U.S. Online Shopping Payment Method Preference 2018."
Choo, K.-K. R. 2011. "The Cyber Threat Landscape: Challenges and Future Research Directions," *Computers & security* (30:8), pp. 719-731.
Chryssochoou, X. 2003. "Studying Identity in Social Psychology: Some Thoughts on the Definition of Identity and Its Relation to Action," *Journal of language Politics* (2:2), pp. 225-241.

Clarke, R. V. G., and Felson, M. 1993. *Routine Activity and Rational Choice* New Brunswick, NJ: Transaction Publishers.

Claub, S., and Kohntopp, M. 2001. "Identity Management and Its Support of Multilateral Security," *Computer networks (Amsterdam, Netherlands : 1999)* (37:2), pp. 205-219.

Clay-Warner, J., Bunch, J. M., and McMahon-Howard, J. 2016. "Differential Vulnerability: Disentangling the Effects of State Dependence and Population Heterogeneity on Repeat Victimization," *Criminal Justice and Behavior* (43:10), pp. 1406-1429.

Cohen, L. E., and Felson, M. 1979. "Social Change and Crime Rate Trends: A Routine Activity Approach," *American sociological review*), pp. 588-608.

Cohen, L. E., Felson, M., and Land, K. C. 1980. "Property Crime Rates in the United States: A Macrodynamic Analysis, 1947-1977; with Ex Ante Forecasts for the Mid- 1980," *The American journal of sociology* (86:1), pp. 90-118.

Cook, P. J. 1986. "The Demand and Supply of Criminal Opportunities," *Crime and justice (Chicago, Ill.)* (7), pp. 1-27.

Copes, H. 1999. "Routine Activities and Motor Vehicle Theft: A Crime Specific Approach," *Journal of crime & justice* (22:2), pp. 125-146.

Copes, H., Kerley, K. R., Huff, R., and Kane, J. 2010. "Differentiating Identity Theft: An Exploratory Study of Victims Using a National Victimization Survey," *Journal of Criminal Justice* (38:5), pp. 1045-1052.

Craig, K., Thatcher, J. B., and Grover, V. 2019. "The It Identity Threat: A Conceptual Definition and Operational Measure," *Journal of Management Information Systems* (36:1), pp. 259-288.

Cram, W. A., Brohman, K., and Gallupe, R. B. 2016a. "Information Systems Control: A Review and Framework for Emerging Information Systems Processes," *Journal of the Association for Information Systems* (17:4), pp. 216-266.

Cram, W. A., Brohman, M. K., Chan, Y. E., and Gallupe, R. B. 2016b. "Information Systems Control Alignment: Complementary and Conflicting Systems Development Controls," *Information & Management* (53:2), pp. 183-196.

Deborah, J. M., and Bernard, J. J. 1989. "Information Processing from Advertisements: Toward an Integrative Framework," *Journal of marketing* (53:4), pp. 1-23.

Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F., and Holz, T. 2019. "We Value Your Privacy ... Now Take Some Cookies: Measuring the Gdpr's Impact on Web Privacy," *Informatik-Spektrum* (42:5), pp. 345-346.

Dehejia, R. H., and Wahba, S. 2002. "Propensity Score-Matching Methods for Nonexperimental Causal Studies," *Review of Economics and Statistics* (84:1), pp. 151-161.

DiPrete, T. A., and Gangl, M. 2004. "Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments," *Sociological Methodology* (34:1), pp. 271-310.

Downing, C., Howard, E. H., Goodwin, C., and Geller, E. S. 2016. "Preventing the Threat of Credit-Card Fraud: Factors Influencing Cashiers' Identification-Checking Behavior," *Journal of prevention & intervention in the community* (44:3), pp. 177-185.

Duck, S., and McMahan, D. T. 2021. *Communication in Everyday Life : A Survey of Communication*. Thousand Oaks, Calif.: Sage Publications, Inc.

Eccles, J. 2009. "Who Am I and What Am I Going to Do with My Life? Personal and Collective Identities as Motivators of Action," *Educational psychologist* (44:2), pp. 78-89.

Eck, J., and Clarke, R. 2003a. "Police Problems: The Complexity of Problem Theory, Research and Evaluation," *Crime prevention studies* (15), pp. 79-114.

Eck, J. E., and Clarke, R. V. 2003b. "Classifying Common Police Problems: A Routine Activity Approach," in *Crime Prevention Studies*, M.J. Smith and D.B. Cornish (eds.). Monsey, NY: Criminal Justice Press, pp. 7-39.

Eisenstein, E. M. 2008. "Identity Theft: An Exploratory Study with Implications for Marketers," *Journal of Business Research* (61:11), pp. 1160-1172.

EPIC. 2018. "Equifax Data Breach." Retrieved July 8th, 2021, from https://epic.org/privacy/data-breach/equifax/

Erikson, E. H. 1968. *Identity, Youth, and Crisis*, ([1st ed.] ed.). New York: W.W. Norton.

FBI. 2019. "2018 Crime Statistics." *Uniform Crime Reporting Program* Retrieved April 20, 2020, from https://ucr.fbi.gov/crime-in-the-u.s/2018/crime-in-the-u.s.-2018/topic-pages/property-crime

FederalRegister. 2003. "Customer Identification Programs for Banks, Savings Associations, Credit Unions and Certain Non-Federally Regulated Banks," F.R.R.a. Regulations (ed.).

Felson, M. 2002. *Crime and Everyday Life : Insights and Implications for Society*. London: Sage Publications.

Gabriel, B., and Dipankar, C. 1983. "Information Accessibility as a Moderator of Consumer Choice," *The Journal of consumer research* (10:1), pp. 1-14.

Ghorbani, M., Liao, Y., Çayköylü, S., and Chand, M. 2013. "Guilt, Shame, and Reparative Behavior: The Effect of Psychological Proximity," *Journal of business ethics* (114:2), pp. 311-323.

Glynos, D., Kotzanikolaou, P., and Douligeris, C. 2005. "Preventing Impersonation Attacks in Manet with Multi-Factor Authentication." IEEE, pp. 59-64.

Goel, R. K. 2019. "Identity Theft in the Internet Age: Evidence from the U.S. States," *Managerial and Decision Economics* (40:2), pp. 169-175.

Golladay, K., and Holtfreter, K. 2017. "The Consequences of Identity Theft Victimization: An Examination of Emotional and Physical Health Outcomes," *Victims & Offenders* (12:5), pp. 741-760.

Gopal, R. D., and Sanders, G. L. 1992. "The Effect of Preventive and Deterrent Software Piracy Strategies on Producer Profits," *ICIS 1992 Proceedings. 9*, pp. 161-170.

Gopal, R. D., and Sanders, G. L. 1997. "Preventive and Deterrent Controls for Software Piracy," *Journal of Management Information Systems* (13:4), pp. 29-47.

Groves, R. M. 2004. *Survey Errors and Survey Costs / Robert M. Groves*. Hoboken, N.J: Wiley-Interscience.

Hamad, A. A. A., Petri, I., Rezgui, Y., and Kwan, A. 2017. "Towards the Innovation of an Integrated 'One-Stop-Shop' Online Services Utility Management: Exploring Customer' Technology Acceptance," *Sustainable cities and society* (34), pp. 126-143.

Harrell, E. 2019. "Victims of Identity Theft, 2016: Bulletin," 1-29/NCJ 251147, US Department of Justice, Office of Justice Programs, Bureau of Justice.

Heinze, G., and Schemper, M. 2002. "A Solution to the Problem of Separation in Logistic Regression," *Statistics in Medicine* (21:16), pp. 2409-2419.

Henson, B., Reyns, B. W., and Fisher, B. S. 2011. "Security in the 21st Century: Examining the Link between Online Social Network Activity, Privacy, and Interpersonal Victimization," *Criminal Justice Review* (36:3), pp. 253-268.

Hill, K. 2012. "How Target Figured out a Teen Girl Was Pregnant before Her Father Did." Retrieved July 8th, 2021

Hille, P., Walsh, G., and Cleveland, M. 2015. "Consumer Fear of Online Identity Theft: Scale Development and Validation," *Journal of Interactive Marketing* (30:2), pp. 1–19.

Hindelang, M. J. 1978. *Victims of Personal Crimes : An Empirical Foundation for a Theory of Personal Victimization / Michael J. Hindelang, Michael R. Gottfredson, James Garofalo*. Cambridge, Mass: Ballinger Pub. Co.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis* (15:3), pp. 199-236.

Holtfreter, K., Reisig, M. D., and Pratt, T. C. 2008. "Low Self-Control, Routine Activities, and Fraud Victimization," *Criminology* (46:1), pp. 189-220.

Hoofnagle, C. J. 2005. "Putting Identity Theft on Ice: Freezing Credit Reports to Prevent Lending to Impostors," in *Securing Privacy in the Internet Age*, A. Chander, L. Gelman and M.J. Radin (eds.). Palo Alto: Stanford University Press, pp. 207-220.

Iacus, S. M., King, G., and Porro, G. 2012. "Causal Inference without Balance Checking: Coarsened Exact Matching," *Political analysis* (20:1), pp. 1-24.

Imbens, G. W., and Rubin, D. B. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. New York, NY: Cambridge University Press.

ITADA. 1998. "Identity Theft Assumption and Deterrence Act of 1998 (Us)," *Public Law*), pp. 105-318.

ITRC. 2020. "Data Breach Report: Are Consumers at Less Risk?," Identity Theft Resource Center.

Iyer, A., and Jetten, J. 2011. "What's Left Behind: Identity Continuity Moderates the Effect of Nostalgia on Well-Being and Life Choices," *Journal of personality and social psychology* (101:1), pp. 94-108.

J.P.Morgan. 2021. "E-Commerce Payments Trends: United States." Retrieved July 8th, 2021, from https://www.jpmorgan.com/merchant-services/insights/reports/united-states

Jansen, J., and Van Schaik, P. 2018. "Testing a Model of Precautionary Online Behaviour: The Case of Online Banking," *Computers in Human Behavior* (87), pp. 371-383.

Jeffrey, C. M. 1988. "Miscellanea, Self/Proxy Response Status and Survey Response Quality, a Review of the Literature," *Journal of official statistics* (4:2), p. 155.

Jung, J., Bapna, R., Ramaprasad, J., and Umyarov, A. 2019. "Love Unshackled: Identifying the Effect of Mobile App Adoption in Online Dating," *MIS Quarterly* (43), pp. 47-72.

Kahn, C. M., and Roberds, W. 2008. "Credit and Identity Theft," *Journal of Monetary Economics* (55:2), pp. 251-264.

Kaiser, H. F. 1974. "An Index of Factorial Simplicity," *Psychometrika* (39:1), pp. 31-36.

Kalvet, T., Tiits, M., and Ubakivi-Hadachi, P. 2019. "Risks and Societal Implications of Identity Theft," A. Chugunov, Y. Misnikov, E. Roshchin and D. Trutnev (eds.). Cham: Springer International Publishing, pp. 67-81.

Kim, A. Y., and Kim, T. S. 2016. "Factors Influencing the Intention to Adopt Identity Theft Protection Services: Severity Vs Vulnerability," in: *PACIS 2016 Proceedings. 68*.

Kumar, N., Qiu, L., and Kumar, S. 2018. "Exit, Voice, and Response on Digital Platforms: An Empirical Investigation of Online Management Response Strategies," *Information Systems Research* (29:4), pp. 849-870.

Kwon, J., and Johnson, M. E. 2014. "Proactive Versus Reactive Security Investments in the Healthcare Sector," *MIS Quarterly* (38:2), pp. 451-472.

Kwon, J., and Johnson, M. E. 2018. "Meaningful Healthcare Security: Does "Meaningful-Use" Attestation Improve Information Security Performance?," *MIS Quarterly* (42:4), pp. 1043-1067.

Lai, F. J., Li, D. H., and Hsieh, C. T. 2012. "Fighting Identity Theft: The Coping Perspective," *Decision Support Systems* (52:2), pp. 353-363.

LaPiedra, J. R. 2014. *Identity Lockdown : Your Step-by-Step Guide to Identity Theft Protection*. Middletown, DE: Lulu Publishing Services.

Lauritsen, J. L. 2001. "The Social Ecology of Violent Victimization: Individual and Contextual Effects in the Ncvs," *Journal of quantitative criminology* (17:1), pp. 3-32.

Leuven, E., and Sianesi, B. 2003. "Psmatch2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing." *Statistical Software Components S432001, Boston College, Chestnut Hill, MA* Retrieved March 10, 2020, from http://ideas.repec.org/c/boc/bocode/s432001.html

Li, Y., Yazdanmehr, A., Wang, J., and Rao, H. R. 2019. "Responding to Identity Theft: A Victimization Perspective," *Decision Support Systems* (121), pp. 13-24.

Lorio, P. J. 2017. "Access Denied: Data Breach Litigation, Article Iii Standing, and a Proposed Statutory Solution," *Columbia Journal of Law and Social Problems* (51:1), pp. 79-128.

Ma, M., and Agarwal, R. 2007. "Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities," *Information systems research* (18:1), pp. 42-67.

Magnus, L., and Steven, R. 2016. "Crime, the Criminal Justice System, and Socioeconomic Inequality," *The Journal of economic perspectives* (30:2), pp. 103-126.

Mank, B. C. 2017. "Data Breaches, Identity Theft, and Article Iii Standing: Will the Supreme Court Resolve the Split in the Circuits?," *Notre Dame Law Review* (92:3), pp. 1323-1367.

Mann, C. 2015. "Information Lost: Will the "Paradise" That Information Promises, to Both Consumer and Firm, Be "Lost" on Account of Data Breaches? The Epic Is Playing Out," in *Economic Analysis of the Digital Economy*. National Bureau of Economic Research, pp. 309-351.

Marchini, K., and Pascual, A. 2019. "2019 Identity Fraud Study: Fraudsters Seek New Targets and Victims Bear the Brunt," Javelin Strategy & Research.

Markowitz, S. 2005. "Alcohol, Drugs and Violent Crime," *International review of law and economics* (25:1), pp. 20-44.

Merriam-Webster, n. d. 2020. "Identity." *In Merriam-Webster.com dictionary.* Retrieved June 13, 2020, from https://www.merriam-webster.com/dictionary/identity

Miller-Osborn, J. 2017. "Credential-Based Attacks: Exposing the Ecosystem and Motives Behind Credential Phishing, Theft and Abuse." Retrieved December 5, 2019, from https://www.paloaltonetworks.com/resources/research/unit-42-credential-based-attacks

Milne, G. R. 2003. "How Well Do Consumers Protect Themselves from Identity Theft?," *Journal of Consumer Affairs* (37:2), pp. 388-402.

Molina-Morales, F. X., García-Villaverde, P. M., and Parra-Requena, G. 2014. "Geographical and Cognitive Proximity Effects on Innovation Performance in Smes: A Way through Knowledge Acquisition," *International entrepreneurship and management journal* (10:2), pp. 231-251.

Monge, P. R., Rothman, L. W., Eisenberg, E. M., Miller, K. I., and Kirste, K. K. 1985. "The Dynamics of Organizational Proximity," *Management science* (31:9), pp. 1129-1141.

Moodysson, J., and Jonsson, O. 2007. "Knowledge Collaboration and Proximity: The Spatial Organization of Biotech Innovation Projects," *European urban and regional studies* (14:2), pp. 115-131.

Moore, T., Clayton, R., and Anderson, R. 2009. "The Economics of Online Crime," *Journal of Economic Perspectives* (23:3), pp. 3-20.

Mustaine, E. E., and Tewksbury, R. 1998. "Predicting Risks of Larceny Theft Victimization: A Routine Activity Analysis Using Refined Lifestyle Measures," *Criminology (Beverly Hills)* (36:4), pp. 829-858.

Navarro, J. C., and Higgins, G. E. 2017. "Familial Identity Theft," *American Journal of Criminal Justice* (42:1), pp. 218-230.

Newman, G. 2004a. "Identity Theft. Problem-Oriented Guides for Police No. 25. Department of Justice, Cops," *Center for Problem-Oriented Policing. Retrieve from http://www. popcenter. org/Problems/problemidentity stheft. htm*).

Newman, G. R. 2004b. *Identity Theft*. Citeseer.

Newman, G. R., and McNally, M. M. 2005a. "Identity Theft Literature Review." *National Institute of Justice Focus Group Meeting* Retrieved April 20, 2020, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.216.6852&rep=rep1&type=pdf

Newman, G. R., and McNally, M. M. 2005b. "Identity Theft Literature Review,").

Norbert, S. 1995. "What Respondents Learn from Questionnaires: The Survey Interview and the Logic of Conversation," *International statistical review* (63:2), pp. 153-168.

Ogbanufe, O., and Pavur, R. 2016. "Going through the "Emotions": Identity Protective Responses," *WISP 2016 Proceedings.1*.

OIG. 2020. "Deploying Unemployment Insurance Benefits Expeditiously While Reducing Improper Payments," United States Department of Labor Office of Inspector General.

PCIDSS. 2018. "Payment Card Industry Data Security Standard Requirements and Security Assessment Procedures V3.2.1," P.S.S. Council (ed.).

Petriglieri, J. L. 2011. "Under Threat: Responses to and the Consequences of Threats to Individuals' Identities," *The Academy of Management review* (36:4), pp. 641-662.

Poster, M. 2007. "The Secret Self: The Case of Identity Theft," *Cultural studies (London, England)* (21:1), pp. 118-140.

Pratt, T. C., Holtfreter, K., and Reisig, M. D. 2010. "Routine Online Activity and Internet Fraud Targeting: Extending the Generality of Routine Activity Theory," *Journal of Research in Crime and Delinquency* (47:3), pp. 267-296.

Pregibon, D. 1981. "Logistic Regression Diagnostics," *The Annals of Statistics* (9:4), pp. 705-724.

Reyns, B. W. 2013. "Online Routines and Identity Theft Victimization: Further Expanding Routine Activity Theory Beyond Direct-Contact Offenses," *The journal of research in crime and delinquency* (50:2), pp. 216-238.

Reyns, B. W., and Henson, B. 2016. "The Thief with a Thousand Faces and the Victim with None: Identifying Determinants for Online Identity Theft Victimization with Routine Activity Theory," *International Journal of Offender Therapy and Comparative Criminology* (60:10), pp. 1119-1139.

Reyns, B. W., Henson, B., and Fisher, B. S. 2011. "Being Pursued Online: Applying Cyberlifestyle–Routine Activities Theory to Cyberstalking Victimization," *Criminal justice and behavior* (38:11), pp. 1149-1169.

Reyns, B. W., and Randa, R. 2017. "Victim Reporting Behaviors Following Identity Theft Victimization: Results from the National Crime Victimization Survey," *Crime & Delinquency* (63:7), pp. 814-838.

Roberds, W., and Schreft, S. L. 2009. "Data Breaches and Identity Theft," *Journal of Monetary Economics* (56:7), pp. 918-929.

Romanosky, S., Telang, R., and Acquisti, A. 2011. "Do Data Breach Disclosure Laws Reduce Identity Theft?," *Journal of Policy Analysis and Management* (30:2), pp. 256-286.

Roncek, D. W., and Faggiani, D. 1985. "High Schools and Crime: A Replication," *Sociological quarterly* (26:4), pp. 491-505.

Rosenbaum, P. 2002. *Observational Studies*, (2nd ed.). Springer, New York.

Rosenbaum, P. R., and Rubin, D. B. 1984. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score," *Journal of the American Statistical Association* (79:387), pp. 516-524.

Rosenbaum, P. R., and Rubin, D. B. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score," *The American Statistician* (39:1), pp. 33-38.

Solove, D. J., and Citron, D. K. 2018. "Risk and Anxiety: A Theory of Data-Breach Harms," *Texas Law Review* (96:4), pp. 737-786.

Steinbart, P. J., Keith, M. J., and Babb, J. 2016. "Examining the Continuance of Secure Behavior: A Longitudinal Field Study of Mobile Device Authentication," *Information Systems Research* (27:2), pp. 219-239.

Straub, D. W. 1989. "Validating Instruments in Mis Research," *MIS Quarterly* (13:2), pp. 147-169.

Straub, D. W., and Welke, R. J. 1998. "Coping with Systems Risk: Security Planning Models for Management Decision Making," *MIS Quarterly* (22:4), pp. 441-469.

Stryker, S., and Burke, P. J. 2000. "The Past, Present, and Future of an Identity Theory," *Social psychology quarterly* (63:4), pp. 284-297.

Sudman, S., Schwarz, N., and Bradburn, N. M. 1996. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass Publishers.

Sun, M., and Zhu, F. 2013. "Ad Revenue and Content Commercialization: Evidence from Blogs," *Management Science* (59:10), pp. 2314-2331.

The United States Department of Justice. 2017. "Identity Theft." Retrieved November 18, 2019, from https://www.justice.gov/criminal-fraud/identity-theft/identity-theft-and-identity-fraud

Tseloni, A., Wittebrood, K., Farrell, G., and Pease, K. 2004. "Burglary Victimization in England and Wales, the United States and the Netherlands: A Cross-National Comparative Test of Routine Activities and Lifestyle Theories," *British journal of criminology* (44:1), pp. 66-91.

US Public Law. 1998. "Law 105-318 (1998). 105th Cong. 112 Stat. 3007,(October 30, 1998)," in: *Identity Theft Assumption and Deterrence Act of 1998*.

US Public Law. 2004. "Law 108-275 (2004), 108th Cong. 118 Stat. 831, (July 15, 2004)," in: *Identity Theft Penalty Enhancement Act of 2004*.

van der Meulen, N. S. 2011. *Financial Identity Theft: Context, Challenges and Countermeasures*. The Hague, Netherlands: T.M.C. Asser Press.

Varian, H. 2004. "System Reliability and Free Riding," in *Economics of Information Security,* L.J. Camp and S. Lewis (eds.). Boston: Kluwer Academic Publishers, pp. 1-15.

Velasquez, E., Ferguson, J., and Cullina, M. 2017. "Identity Theft: The Aftermath 2017." *Identity Theft Resource Center* Retrieved March 25, 2020, from https://www.ftc.gov/system/files/documents/public_comments/2017/10/00004-141444.pdf

Virtue, T., and Rainey, J. 2015. "Chapter 6: Information Risk Assessment," in *Hcispp Study Guide*. Waltham, MA: Syngress, pp. 131-166.

Voci, A. 2006. "The Link between Identification and in-Group Favouritism: Effects of Threat to Social Identity and Trust-Related Emotions," *British journal of social psychology* (45:2), pp. 265-284.

von Proff, S. 2016. "The Predominance of Social Proximity for Innovation Collaboration of Sme," Working Papers on Innovation and Space.

Wang, J., Yazdanmehr, A., Li, Y., and Rao, H. R. 2017. "Opting for Identity Theft Protection Services: The Role of Anticipated Distress," *ICIS 2017 Proceedings. 14*.

Whitbourne, S. K., and Connolly, L. A. 1998. *Life in the Middle : Psychological and Social Development in Middle Age / Edited by Sherry L. Willis, James D. Reid*. San Diego: Academic Press.

Williams, M. L. 2016. "Guardians Upon High: An Application of Routine Activities Theory to Online Identity Theft in Europe at the Country and Individual Level," *British journal of criminology* (56:1), pp. 21-48.

Wysopal, C. 2020. "Determining Liability for Security Breaches Isn't Black and White." *In Forbes.com.* Retrieved June 13, 2020, from https://tinyurl.com/y77fdpdg

Xie, M., and Baumer, E. P. 2019. "Neighborhood Immigrant Concentration and Violent Crime Reporting to the Police: A Multilevel Analysis of Data from the National Crime Victimization Survey," *Criminology (Beverly Hills)* (57:2), pp. 237-267.

Yar, M. 2005. "The Novelty of 'Cybercrime': An Assessment in Light of Routine Activity Theory," *European journal of criminology* (2:4), pp. 407-427.

Ylang, N. 2020. "Capable Guardianship against Identity Theft: Demographic Insights Based on a National Sample of Us Adults," *Journal of Financial Crime* (27:1), pp. 130-142.

Young, K. 2020. *The Domains of Identity: A Framework for Understanding Identity Systems in Contemporary Society*. New York, NY: Anthem Press.

Zaiss, J., Nokhbeh Zaeem, R., and Barber, K. S. 2019. "Identity Threat Assessment and Prediction," *Journal of Consumer Affairs* (53:1), pp. 58-70.

Zviran, M., and Erlich, Z. 2006. "Identification and Authentication: Technology and Implementation Issues," *Communications of the Association for Information Systems* (17), p. 4.