Journal of the Association for Information Systems

Volume 25 | Issue 4 Article 2

2024

An Empirical Investigation of The Unintended Consequences of Vulnerability Assessments Leading to Betrayal

Dan Pienta
University of Tennessee, Knoxville, dpienta@utk.edu

Jason Bennett Thatcher
University of Colorado Boulder / University of Manchester, jason.thatcher@colorado.edu

Ryan T. Wright
University of Virginia, rtwright@virginia.edu

Philip L. Roth
Clemson University, rothp@clemson.edu

Recommended Citation

Pienta, Dan; Thatcher, Jason Bennett; Wright, Ryan T.; and Roth, Philip L. (2024) "An Empirical Investigation of The Unintended Consequences of Vulnerability Assessments Leading to Betrayal," *Journal of the Association for Information Systems*, 25(4), 1079-1116.
DOI: 10.17705/1jais.00875
Available at: https://aisel.aisnet.org/jais/vol25/iss4/2

Journal of the Association for Information Systems (2024) **25**(4), 1079-1116
doi: 10.17705/1jais.00875

RESEARCH ARTICLE

ISSN 1536-9323

An Empirical Investigation of The Unintended Consequences of Vulnerability Assessments Leading to Betrayal

Daniel Pienta,[1] Jason Bennett Thatcher,[2] Ryan T. Wright,[3] Philip L. Roth[4]

[1]University of Tennessee, Knoxville, USA, dpienta@utk.edu
[2]University of Colorado Boulder, USA / University of Manchester, UK, jason.thatcher@colorado.edu
[3]University of Virginia, USA, rtwright@virginia.edu
[4]Clemson University, USA, rothp@clemson.edu

Abstract

When cybersecurity units conduct vulnerability assessments to evaluate the security of organizations, they can have unintended consequences for employees. Although cybersecurity personnel may view tactics such as fake phishing attacks and email scanning as protective measures, employees may view them as threats because being singled out as a security risk can harm their standing in the organization. To understand the implications of vulnerability assessments, we examine how organizations' use of different tactics to identify user vulnerabilities can lead employees to feel betrayed by the cybersecurity unit, resulting in negative cybersecurity outcomes. Drawing on the theory of betrayal aversion, we develop a model that shows that when employees perceive these tactics as harmful, they can lead to an affective state of cybersecurity betrayal, resulting in a damaged relationship with the cybersecurity unit. In collaboration with an organization's cybersecurity unit, we evaluated our model using an experimental vignette survey, post hoc interviews, and a cross-sectional survey with two samples (i.e., employees in the organization and employees from a panel). We found that when organizations conduct vulnerability assessments to enhance cybersecurity, they often induce an affective state of betrayal and increase employees' active resistance to cybersecurity (i.e., abandonment, avoidance, and sabotage of cybersecurity policies, technologies, and units). The paper concludes with implications for research and practice that explain the unintended consequences of vulnerability assessment and betrayal.

Keywords: Betrayal, Active Resistance, Cybersecurity, Vulnerability Assessment, Unintended Consequences

Hillol Bala was the accepting senior editor. This research article was submitted on March 4, 2022, and underwent three revisions.

1 Introduction

Organizations rely on layered cybersecurity defenses to identify, protect, detect, respond to, and recover from potential cybersecurity attacks (NIST, 2018). Often, layered defenses focus on identifying threats posed by people, processes, and technology (Pienta et al., 2020), which regulators recommend should be continuously audited, monitored, and tested as part of risk assessment programs (Ross, 2012). These tactics are referred to as vulnerability assessments and identify weaknesses by auditing, monitoring, or testing (e.g., reviewing logs, scanning all network devices and programs, and penetration testing) the organization's technology, processes, and people.

Vulnerability assessments include active threat detection, such as phishing simulations and scanning of computers for data misuse, as well as the covert monitoring of email, social media, network traffic, and even productivity tools such as Word. The results of

these assessments allow the organization to strengthen its cybersecurity defenses by reducing the risk associated with vulnerabilities. These assessments are becoming increasingly institutionalized, with standard-setting bodies—such as the PCI Security Standards Council, the National Institute of Standards and Technology (NIST), and the International Organization for Standards (ISO)—requiring their use to maintain compliance with industry standards for technical and operational cybersecurity components (PCI, 2017). These industry standards explicitly specify vulnerability assessments, such as activity testing and employee monitoring, as best practices (e.g., NIST, 2021).

However, even if they are not (formally) identified as a problem, employees often react negatively if they feel that a vulnerability assessment or its results could embarrass them, damage their performance reviews, or lead to disciplinary action (Kelley et al., 2012; Krebs, 2019; Wright & Thatcher, 2021). These feelings are rooted in how organizations treat employees who pose cybersecurity risks. More than 42% of organizations discipline employees for cybersecurity mistakes, such as clicking on a simulated phishing link. For example, 63% of cybersecurity training errors are reported to supervisors, 15% of these organizations name and shame employees, 33% reduce access privileges, and 17% block employees from resources until mandatory training is completed (Blythe & Collins, 2022). As seen in the fallout from phishing simulations, when employees view a person or organization as a source of harm and betrayal, they often resist future efforts from that source (Craig et al., 2019) because it jeopardizes their position in the organization (Martinko et al., 1996; Rivard & Lapointe, 2012). The unintended result may be that vulnerability assessments damage the relationship between employees and the cybersecurity unit by making them feel harmed by the cybersecurity unit's actions.

When employees potentially feel harmed by cybersecurity, they may feel betrayed, which can have negative consequences for an organization's cybersecurity. One such example occurred in 2020, when GoDaddy sent a phishing test from happyholiday@godaddy.com to more than 500 employees, claiming they had received a holiday bonus (see Figure 1). Two days later, the company's chief security officer wrote to those who added information to the form, saying: "You're getting this email because you failed our recent phishing test. You will need to retake the security awareness social engineering training" (Longhi, 2020).

Other companies, such as Tribune Publishing, which owns major newspapers in the United States, have also had phishing simulations go awry by offering fake bonuses to employees. The cybersecurity units of both GoDaddy and Tribune Publishing received a backlash from employees (Wagner, 2020). One security expert stated: "These types of simulated phishing emails can leave a terrible taste in employees' mouths and even cause employees to question your company's values. All it takes is one lousy phishing test to destroy trust and create a company culture of doubt" (Anders, 2022, p. 1). In addition, these phishing simulations resulted in a public backlash that likely tarnished the image and reputation of both organizations (Wagner, 2020).

Figure 1. GoDaddy Simulated Email, December 14, 2020 (Longhi, 2020)

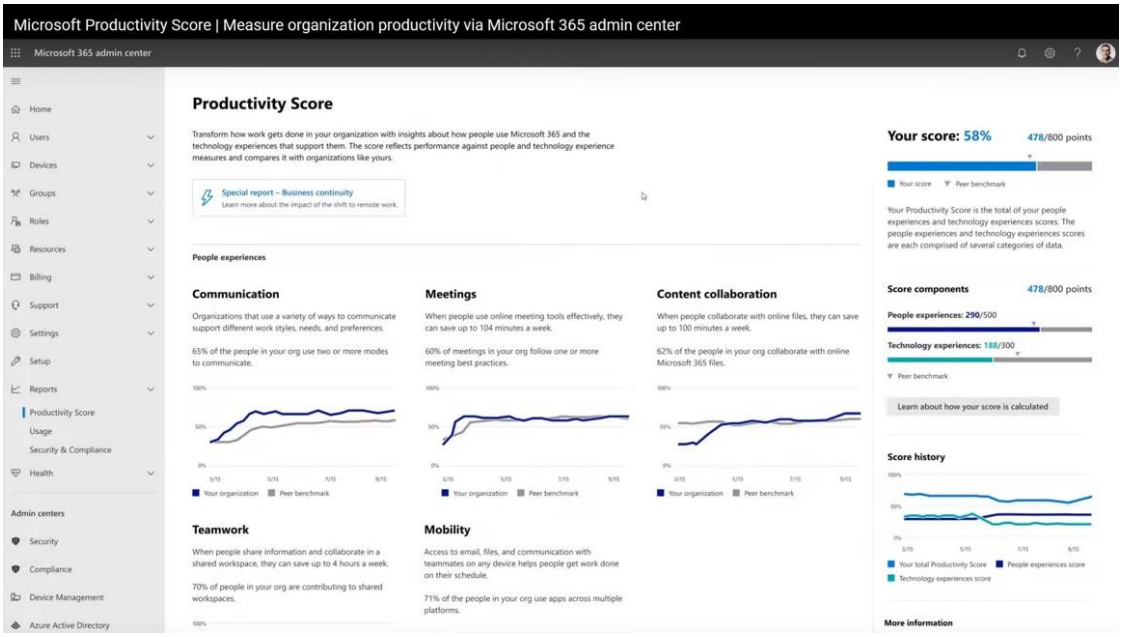


Figure 2. Example of Microsoft Productivity Score (Microsoft, 2023)

In addition to poorly executed phishing simulations, we suspect that many tactics used to assess vulnerabilities can raise employee concerns. For example, employees are increasingly concerned about organizations monitoring email and personal social media accounts for cybersecurity purposes. A Gartner study found that of the 239 large organizations surveyed, nearly 50% monitored emails and social media (Accenture, 2019). Employees reported being "incredibly stressed out" by this type of monitoring (Harwell, 2020).

Moreover, even well-designed cybersecurity tools may evoke concern among employees. Security and monitoring tools used to be available primarily from fringe software vendors but are now being integrated into products such as Office 365. For example, in October 2020, Microsoft announced a tool that provides a productivity score for remote workers (Microsoft, 2020) by ingesting emails, meetings, collaborative documents (such as Word), and other Microsoft tools to estimate an employee's productivity score (see quote below and Figure 2). There has been significant backlash following the release of this tool (Cater & Heikkile, 2021), in part because of a lack of evidence regarding the effectiveness of such tools on actual security and productivity (Mims, 2022).

> *Microsoft 365 includes a broad set of information protection and compliance capabilities. Together with Microsoft's productivity tools, these capabilities are designed to help organizations collaborate in real time while adhering to stringent regulatory compliance frameworks.* (Microsoft, 2023)

If employees view vulnerability assessments, such as phishing tests, monitoring tools, and scans, as potentially harmful, especially if they believe that testing, monitoring, and auditing will result in negative outcomes such as mandatory security training, disciplinary action, a negative performance review, or even termination, then feelings of harm could translate into feelings of betrayal. In turn, feelings of betrayal can lead users to avoid, abandon, or even sabotage the cybersecurity unit, its technologies, and its policies. To explore the possible unintended consequences of vulnerability assessments and betrayal, we apply the theory of betrayal and ask the following two questions:

RQ1: Do employees feel that vulnerability assessments by cybersecurity units are a source of harm?

RQ2: If so, does this lead to a state of cybersecurity betrayal, and what are its consequences?

By exploring the link between vulnerability assessments, cybersecurity harm, and betrayal and its unintended consequences, this research contributes to the cybersecurity literature in several ways. First, our cybersecurity harm and betrayal framework explains why employees resist and work around the cybersecurity unit's technologies and policies. In doing so, our study helps answer why such tactics, which cybersecurity professionals view as protective, are perceived by employees as potentially harmful. Second, our empirical analysis shows how vulnerability assessments lead to feelings of harm and betrayal, which, in turn, lead to active resistance. By extending the research on intentional noncompliant computer misuse (D'Arcy et al., 2014; Willison et al., 2018), our findings provide practitioners with insights into how vulnerability assessments may elicit negative employee reactions and behaviors and thus suggest that researchers and practice partners should consider how to implement such measures in nonharmful ways.

Table 1. Theoretical Constructs and Definitions

Construct	Definition
Vulnerability assessment	The process and tactics associated with identifying, quantifying, and prioritizing cybersecurity weaknesses in an organization
Cybersecurity harm	A belief that cybersecurity individuals, units, or technologies have caused malice to a user
Cybersecurity betrayal	An affective psychological state that results from believing that the cybersecurity unit has violated the core expectations of a protective relationship (Elangovan & Shapiro, 1998; Grégoire & Fisher, 2008; Koehler & Gershoff, 2003)
Betrayal aversion	Behaviors that risk diminishing protection to reduce the mere possibility of betrayal (Koehler & Gershoff, 2003)
Agent of protection	An individual, unit, or technology that is perceived as safeguarding organizations, users, technology, and data
Agent of harm	An individual, unit, or technology that is perceived as enacting malice even if it is designed to protect users or organizations

2 Theoretical Basis

The theory of betrayal aversion focuses on the sources of betrayal and its consequences. This theory suggests that when the agents of protection, such as cybersecurity units, threaten even the possibility of betrayal, people will take action to protect themselves, as they now view the agent of protection as a source of possible harm (Koehler & Gershoff, 2003). In a series of experiments, Koehler and Gershoff (2003) found that when people view agents of protection as a source of harm, they engage in betrayal aversion or behaviors that compromise protection to reduce the mere possibility of betrayal. Their findings showed that individuals attempt to minimize the likelihood of betrayal by making suboptimal protective decisions (e.g., forgoing vaccinations, turning off airbags in a car, or refusing police assistance in an emergency). Often, these decisions are based on violating the underlying *pivotal* (i.e., *core*) expectations of the relationship that involve protection rather than potential harm (i.e., an individual will be harmed by an agent of protection and can be exploited because of the perceived asymmetry in the relationship (e.g., the agent of protection is less vulnerable than those they protect). Table 1 lists the theoretical constructs and definitions.

In most organizations, employees expect the actions of the cybersecurity unit to protect not only the content of their electronic communications but also the data stored on devices connected to organizational assets. For example, employees expect the cybersecurity unit to provide the training and tools necessary to protect personal and organizational devices used for work-related purposes (see Jensen et al., 2021). Although employees see themselves as the primary beneficiaries of cybersecurity, they often forget that employee error or negligence is a source of data breaches. Therefore, even when taking actions to protect employees, cybersecurity units must view employees as potential vulnerabilities (Balozian et al., 2023; Burns et al., 2022; Crossler et al., 2013). This view is echoed by security solution vendors, who often characterize people as the core security problem:

> *There's no denying that humans are the weakest link in cybersecurity. No matter how strong your technical defenses, such as firewall, IPS, or IDS, are, they can always be circumvented by a determined attacker if they can find a way to trick or coerce a member of your staff into giving them access.* (Malik, 2023)

This contradiction between employees seeing themselves as the beneficiaries of protective measures and cybersecurity units also seeing them as a source of vulnerability creates a tension that can lead employees to feel betrayed by the actions of cybersecurity units. For example, when employees are included in vulnerability assessments, they may feel that their expectations of protection have been violated, even if a cybersecurity unit's tactics are effective at protecting firm data, especially if they are punished for a data breach (Cram et al., 2019). Betrayal stems from employees feeling harm or from the sense of the potential for loss or negative consequences from the actions of a trusted entity (Rachman, 2010).

Betrayal is a relevant theoretical foundation for understanding how people respond to the harmful effects of unanticipated shocks or events, such as protection failures (Koehler & Gershoff, 2003), institutional failures (Smith & Freyd, 2014), and service failures (Elangovan & Shapiro, 1998). Betrayal can lead to the erosion of employer-employee relationships in the workplace (Reina & Reina, 2006). Studies have found that betrayal damages relationships, creates negative attitudes, and leads to counterproductive behavioral consequences (Coyle-Shapiro et al., 2019), which warrants further study in the context of cybersecurity. For example, cybersecurity research has shown that users feel betrayed when their reputations are damaged by a

training exercise that catches them violating rules about data sharing or password protection (Blythe & Collins, 2022). Thus, theories of betrayal in the workplace and related evidence suggest that such actions by the cybersecurity unit may induce a state of betrayal, which leads users to engage in behaviors that undermine the organization.

Although most research has examined betrayal in the workplace as a result of unexpected shocks or jarring events (e.g., institutions covering up scandals, products causing deaths, etc.), organizations also need to understand how betrayal can result from routine organizational activities, including cybersecurity vulnerability assessments. This is important because current cybersecurity research shows that employees engage in maladaptive responses (that appear to be to the organization) to routine cybersecurity measures (Balozian et al., 2023).

Hence, we need to develop an understanding of how cybersecurity betrayal leads to active cybersecurity resistance behaviors, which weaken cybersecurity. In this research, we note there are at least three contexts for cybersecurity research to explore: the organization and its internal and third parties (Table 2). Second, this study begins to explore the consequences of betrayal in the organizational context through a protective relationship between the cybersecurity unit and employees. Third, internal stakeholders, such as supervisors and peers, may cause employees to feel betrayed and engage in active resistance. Finally, third parties outside the organization, such as suppliers, outsourcers, and contractors, can lead to feelings of betrayal. Fundamentally, however, as shown in Table 2, existing research in various management disciplines has studied betrayal, demonstrating negative emotions and subsequent behavioral responses that cybersecurity research can build upon.

Table 2. Consequences of Betrayal Violations (Coyle-Shapiro et al., 2019)

Context	Attitude	Behavior	Cybersecurity active resistance behavior
Organization	• Affective organization commitment - (Restubog et al., 2006) • Turnover intentions + (Orvis et al., 2008) • Organizational trust - (Robinson & Wolfe Morrison, 2000) • Job satisfaction - (Conway et al., 2011) • Perceived organizational support (Coyle-Shapiro & Conway, 2005) • Organizational cynicism + (Johnson & O'Leary-Kelly, 2003) • Organizational identification - (Zagenczyk et al., 2011)	• Performance - (Costa & Neves, 2017) • Organizational citizenship behavior (OCB) - (Restubog et al., 2009) • Voice + (Ng et al., 2014) • Workplace deviance + (Bordia et al., 2008) • Absenteeism + (Deery et al., 2006) • Turnover + (Karagonlar et al., 2016)	• Not following compliance requirements + • Not reporting phishing emails for collective defense + • Verbally disparaging cybersecurity + • Using shadow systems (e.g., Dropbox, Gmail) + • Skipping training +
Internal third parties (e.g., supervisors, coworkers)	• Leader-member exchange - (Costa & Neves, 2017)	• Interpersonal deviance + (Bordia et al., 2008) • Interpersonal OCB - (Rosen et al., 2009) • Interpersonal harming toward coworkers + (Deng et al., 2018)	• Purposefully providing others' login credentials + • Not securing shared computers + • Installing unauthorized software on others' computers +
External third parties (e.g., customers, suppliers, friends, family)	• Union commitment - (Turnley et al., 2004) • Public sector commitment - (Conway et al., 2014)	• OCB oriented to public service users - (Conway et al., 2014) • OCB toward the customer - (Bordia et al., 2010) • Decision-making vigilance for clients - (Deng et al., 2018) • Work-nonwork conflict + (Gracia et al., 2007) • Work-family conflict + (Jiang et al., 2017)	• Posting sensitive data on an unsecured server + • Not having the correct access or controls on sensitive files + • Using personal devices when traveling internationally when against policy +

Note: "+" denotes a positive relationship; "-" denotes a negative relationship

2.1 Security Initiatives and Outcomes in the Information Security Literature

The cybersecurity literature (Table 3) is filled with studies that provide a rich theoretical understanding of how security education and awareness (SETA) initiatives affect end users in terms of subsequent security behaviors (e.g., compliance, proactive security behaviors, and extra-role behaviors). Although the literature has mostly focused on positive user responses and outcomes (Cram et al., 2019), there is an emerging stream suggesting that SETA can have negative consequences on employee behaviors. The nascent nature of this stream means that there is a limited theoretical understanding of the negative consequences of SETA, which is primarily focused on decreasing breaches. Negative consequences can result from the policy or from an employee's illegal behavior (e.g., negligence or computer abuse). Often, information security research that examines the negative consequences of SETA programs and policies focuses on developing an understanding of how users perceive/respond to the training or policy and how behaviors change in response to the policy (Cram et al., 2019; Moody et al., 2018).

However, there is scant research (to the authors' knowledge, none in academic literature review) on how SETA programs or other actions by the cybersecurity unit may damage or enhance the relationship between employees and those who write policy, develop training, or conduct auditing, monitoring, or testing. In cybersecurity, these tasks are often the responsibility of the cybersecurity unit, and understanding employees' perceptions of the cybersecurity unit may provide a more nuanced understanding of why SETA, policy, and vulnerability assessments work or do not work in certain organizations. Interestingly, as noted in Table 2, in the broader management literature there is a focus on how relationships between people in organizations affect individual behaviors. Therefore, we turn to articulating a research model that explains how breaches in this exchange relationship between employees and cybersecurity units can lead to negative cybersecurity outcomes (Coyle-Shapiro & Conway, 2005).

Table 3. Positive and negative SETA Outcomes in the Cybersecurity Literature

Positive SETA outcomes		
Theory	**Outcome**	**Reference**
Mindfulness	Antiphishing +	(Jensen et al., 2017; Nguyen et al., 2023)
Gamification	Antiphishing +	(Dincelli & Chengalur-Smith, 2020; Jensen et al., 2022; Silic & Lowry, 2020)
Protection motivation/extended parallel processing model/construal level	Compliance + Noncompliance - Antiphishing + Security tool adoption +	(Boss et al., 2015; Chen et al., 2021; Herath & Rao, 2009; Johnston et al., 2016; Johnston et al., 2015; Khern-am-nuai et al., 2023; Mady et al., n.d.; Mattson et al., 2023; Menard et al., 2017; Nehme & George, 2022; Ng et al., 2021; Posey et al., 2011, 2015; Schuetz et al., 2020a, 2020b; Wang et al., 2017; Warkentin et al., 2016; Zahedi et al., 2015)
Deterrence	Compliance +	(Burns et al., 2022; Chen et al., 2012; D'Arcy & Herath, 2011; D'Arcy et al., 2009; Park et al., 2017; Sarkar et al., 2020; Willison et al., 2018)
Proactivity	Compliance intention +	(Boss et al., 2009; Burns et al., 2019)
Rational choice	Compliance +	(Bulgurcu et al., 2010)
Elaboration likelihood	Compliance +	(Puhakainen & Siponen, 2010)
Accountability	Violation -	(Vance et al., 2015)
Safety climate model	Compliance +	(Goo et al., 2014)
Rational choice, psychological contract	Compliance +	(Han et al., 2017)
Social cognitive	Moral disengagement -	(Herath et al., 2018)
Reactance, protection motivation, organizational justice	Compliance intention +	(Hovav & Putri, 2016)
Prospect, protection motivation	Noncompliance -	(Hwang et al., 2017)
Social bond/involvement	Compliance +	(Safa et al., 2016)
Protection motivation, theory of planned behavior	Compliance +	(Moquin & Wakefield, 2016)

Social judgment	Antiphishing +	(Jensen et al., 2020)
Heuristic systematic processing model	Antiphishing +	(Goel et al., 2017)
Instructional design	Antiphishing +	(Dodge Jr et al., 2007; Kumaraguru et al., 2010)
Managerial security risk	Security risk -	(Straub & Welke, 1998)
Security awareness	Data breach - Awareness +	(Jaeger & Eckhardt, 2021; Weixun Li et al., 2023)
Social control theory	Information security policy effectiveness +	(Feng et al., 2019; Hsu et al., 2015)
Social disorganization theory/collective efficacy	Compliance +	(Johnston et al., 2019; Nguyen et al., 2021; Yazdanmehr et al., 2022)
Psychological empowerment	Compliance intention +	(Dhillon et al., 2020)
Neutralization	Violations -	(Barlow et al., 2018; Barlow et al., 2013)
Selective organizational rules violations	Violations -	(Wall et al., 2015)
Technology threat avoidance	Compliance +	(Liang & Xue, 2010)
Compromising actor-network theory	Security awareness +	(Tsohou et al., 2015)
Interest and self-determination theory	Compliance +	(Kam et al., 2022)
Negative SETA outcomes		
Theory	**Outcome**	**Reference**
Security stress/security fatigue	Violation +	(Cram et al., 2021; D'Arcy et al., 2014; Yazdanmehr et al., 2022, 2023)
Neutralization	Violation +	(Siponen & Vance, 2010; Trinkle et al., 2021)
Habit	Compliance -	(Anderson et al., 2016; Jenkins et al., 2016; Vance et al., 2018)
Embedded training and awareness	Training efficacy -	(Caputo et al., 2013)
Fairness, reactance	Resistance +	(Lowry & Moody, 2015; Lowry et al., 2015)
Beliefs, actions, outcomes model	Maladaptive security behavior +	(Balozian et al., 2023)
Composite behavior model	Nonmalicious security violations +	(Guo et al., 2011)
Capabilities	Decision bias +	(Jalali et al., 2019)
Institutional theory	Resistance +	(Hu et al., 2007)

Note: We searched the AIS basket of 10 journals and journal articles identified in Cram et al. (2019) as including SETA. For the journal articles, we searched based on the keywords SETA, Security Education Training and Awareness, Security Training, Security Education, Security Awareness, and Security Compliance Training, Education, and Awareness. We included articles that measured a form/construct of SETA or manipulated/used a form of SETA. "+" denotes a positive relationship; "-" denotes a negative relationship

3 Research Models and Hypotheses

To establish that vulnerability assessments can make employees feel potential harm and betrayal and to assess the implications of this relationship for active cybersecurity resistance, we partnered with the cybersecurity unit of a real organization to conduct a multimethod series of studies. The purpose of our first study was to examine the relationship between vulnerability assessments and harm. This important theoretical linkage must be established (or not) before the subsequent cascading effects of cybersecurity harm can be explored. To examine the relationship between vulnerability assessments and harm, we used an experimental vignette methodology to elicit responses from current employees about whether vulnerability assessments evoke feelings of potential cybersecurity harm. Next, we conducted a survey to understand the unintended consequences of the cascading effect of vulnerability assessment, harm, and betrayal and how they affect employees' active resistance to cybersecurity behaviors. Finally, we conducted post hoc interviews and robustness tests to confirm our research model. Below, we explain our research model (see Figure 3) and describe our studies.

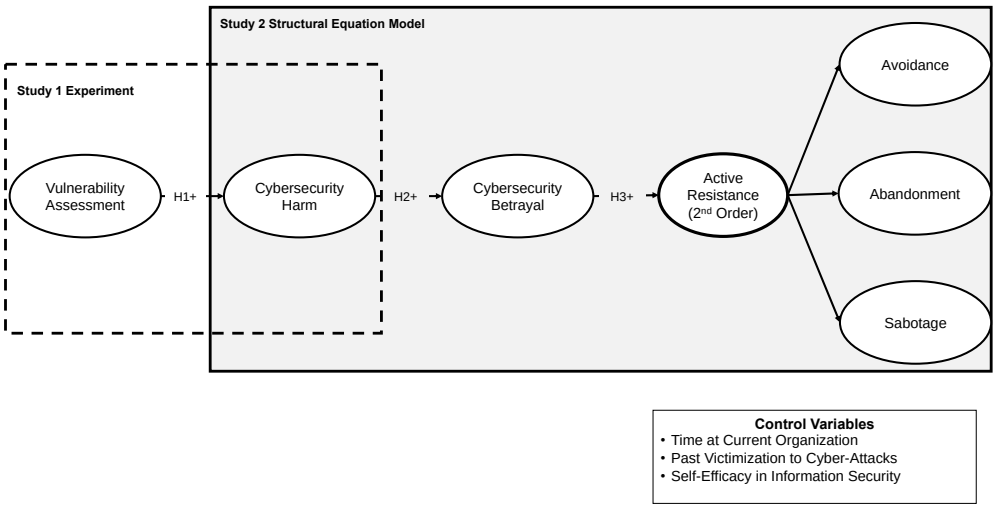


Figure 3. Research Model of Vulnerability Assessments, Harm, and Betrayal

3.1 The Vulnerability Assessment as a Source of User Harm

Vulnerability assessments often mimic cybercriminal techniques. Many of the network, device, and user-oriented tools used to perform these assessments are widely available in open source software (e.g., Kali Linux, Backtrack). Cybersecurity units use fake phishing (e.g., asking for passwords or logins), vulnerability scanning (e.g., requesting unauthorized data movement or exploiting vulnerable devices), and data exfiltration tactics (e.g., stealing passwords from hacked email accounts) to mimic criminals' tactics. For example, penetration tests often attempt to convince employees to click on malicious links or download malicious files that create opportunities to steal confidential information (Wright et al., 2014). When used effectively, these tools can identify vulnerabilities in technology, processes, and people.

Although organizations have the legal right to conduct vulnerability assessments if employees are informed by policy, the organization can still withhold the specifics of the assessment (e.g., time, type of test, etc.) making them no-notice tests. This allows cybersecurity units to conduct vulnerability assessments and decide how and whether to conduct assessments without informing employees of the specifics of the test. Cybersecurity units must balance which actions to take because previous research has shown that many users automatically agree with organizational security and privacy policies (Acquisti & Grossklags, 2005), ignore them (Steinfeld, 2016), or lack the security awareness to understand what they have agreed to (Bulgurcu et al., 2010; Tsohou et al., 2015). Furthermore, empirical evidence in the behavioral cybersecurity literature suggests that employees are more likely to be noncompliant with cybersecurity policies due to a misunderstanding of the actual policies, thus bringing about the need to increase employee compliance (Chen et al., 2021; Cram et al., 2019; Crossler et al., 2017). Often, the misunderstanding of policies and cybersecurity incident alerts is caused by employees perceiving cybersecurity policies as vague, nebulous, complex, changing, or lengthy (Milne & Culnan, 2004; Tsai et al., 2011; Vance et al., 2019).

Cybersecurity units conduct vulnerability assessments in different ways (e.g., how much, if any, of the specifics of the assessment are provided to employees), but these can primarily be categorized as notice or no-notice to employees and passive or active actions (NIST, 2022) (Table 4). A passive action that can be noticed by employees in real time (i.e., notice in Table 4) includes labeling the source of an email as internal or external or blocking access to known malicious websites. Active actions that employees notice in real time include virus scans and threat alerts. Passive actions that are not visible to employees in real time (i.e., no-notice in Table 4) include anomaly detection (such as event correlation) and content filtering. Active actions that are not visible to employees in real time include fake phishing attacks that trick employees into clicking malicious links, downloading malicious files, and sharing confidential information (Wright et al., 2014), and the active scanning of devices and information for vulnerabilities. Noticeable and unnoticeable (i.e., notice and no-notice), real-time, passive vulnerability assessments, such as spam filters and anomaly detection (e.g., standard monitoring), are often accepted by employees and, therefore, may not be perceived as harmful because they protect employees (George, 1996).

Table 4. Vulnerability Assessment Orientation

	Passive	Active
Notice Employees are made aware	• Adding a warning label to all external emails. • Blocking malicious websites on a blacklist.	• Antivirus scans on a workstation. • Warning of a possible phishing message in an email. • Monitoring email.
No-notice Employees are unaware	• Intrusion detection systems that detect anomalous traffic. • Firewalls that block harmful websites, inbound email, and messaging traffic.	**[Our focus]** • Phishing simulations. • Monitoring social media and productivity tools.

Perhaps the most problematic cyber efforts are no-notice, active vulnerability assessments, such as fake phishing attacks, and vulnerability scans that exfiltrate employee data (e.g., logging hosts for auditing) because they can evoke feelings of harm and betrayal in employees (Wright & Thatcher, 2021). For example, cybersecurity units use scanning and protection software on personal devices to protect organizational assets, such as email, access to files, and other proprietary data (Bacudio et al., 2011; Kohnke et al., 2017). When cybersecurity units describe these measures as protecting personal devices, employees can reasonably expect protection. However, when cybersecurity units conduct vulnerability scans on personal devices to detect potential vulnerabilities, this action may violate employees' expectations of protection (Colnago et al., 2018) and create feelings of harm. As practitioners have noted, many average users of technology lack a knowledge and understanding of networks, security, and privacy (Pew Research Center, 2017, 2019) and thus may feel harmed if personal devices are accessed by their employer.

We draw on the theory of betrayal aversion, which suggests that when individuals perceive that an agent of protection (i.e., in this context, the cybersecurity unit) can harm them, they feel betrayed and engage in suboptimal security behaviors. Fundamentally, the theory posits that there is a pivotal expectation in protective relationships that the protector will not harm the protected due to this asymmetric vulnerability. In protective relationships, the individual is far more vulnerable to harm in the sense that the protector (e.g., the cybersecurity unit) is privileged to far more information than the protected (e.g., the employee) (Koehler & Gershoff, 2003). This feeling of asymmetry stems from two sources: (1) access and (2) indetectability (Koehler & Gershoff, 2003). Access is the concept that one party has the "means to exploit the vulnerability in the other" (Koehler & Gershoff 2003, p. 247). In cybersecurity, the individual is far more vulnerable to harm from the cybersecurity unit in the sense that the function is privileged to far more information (e.g., passwords, access rights, identifiers, etc. from work and personal devices). Indetectability is the idea that the protective party believes they have the ability to avoid detection and punishment (Koehler & Gershoff, 2003). The cybersecurity unit can covertly access information about employees for disciplinary purposes, while employees have limited recourse. Relatedly, employees face far more detrimental consequences if they choose to openly undermine the cybersecurity unit. In such asymmetrical protection relationships, individuals may feel that the protector is a source of harm who can betray them, as they are inherently vulnerable in the relationship.

The theory of betrayal aversion suggests that if users are unaware of cybersecurity actions or the depth of vulnerability assessments and learn about them through discipline or word of mouth, they may feel that cybersecurity units could harm their standing in the organization. As mentioned previously, practice is littered with reports of questionable phishing tests gone wrong (e.g., GoDaddy) and questionable monitoring practices upsetting users (e.g., Microsoft's productivity score) and, in this case, it is often the job of cybersecurity units to audit, monitor, and test information systems and employees (Kohnke et al., 2017). Often, employees in the organization are concerned about their performance on these assessments when they are reported to line managers or discover they have made a security mistake. Although these measures are intended to protect people, they can lead to feelings of cybersecurity harm. Indeed, while these best practices may have positive outcomes for cybersecurity, they may also be flawed (the focus of this research), given that the literature has not explicitly provided an understanding of no-notice vulnerability assessments; therefore, we hypothesize:

H1: No-notice vulnerability assessments can lead to perceived cybersecurity harm among users.

3.2 Cybersecurity Betrayal

Recall that we have defined cybersecurity betrayal as the affective psychological state that results from the belief that the cybersecurity unit has violated the core expectations of a protective relationship (Elangovan & Shapiro, 1998; Grégoire & Fisher, 2008; Koehler & Gershoff, 2003). A state of betrayal may occur when employees perceive the cybersecurity unit's actions as a source of harm that elicits the characteristics of betrayal (Table 5).

Table 5. Characteristics of Betrayal

Characteristic	Definition	Cybersecurity example
Voluntary act	"Trustee either lacks the motivation to conform to expectations of the trustor or becomes motivated to violate these expectations" (Elangovan & Shapiro, 1998, p. 550).	The cybersecurity unit begins to intentionally identify and target negligent employees.
Pivotal expectations	"Must be instrumental to the relationship between the trustor and trustee … May be task or value related as long as they are personal expectations and pivotal to the relationship" (Elangovan & Shapiro, 1998, p. 550).	The employee believes that the cybersecurity unit will protect each employee.
Mutually understood expectations	"Both parties must be mutually aware of (even if it is implicit) but need not necessarily accept these expectations as central to the relationship of contract" (Elangovan & Shapiro, 1998, p. 550).	The cybersecurity unit and the individual both perceive that the goal of cybersecurity is to protect the individual employee and the organization's data.
Violation of expectations	"Involves a behavior (an actual violation) rather than just the thought of betraying" (Elangovan & Shapiro, 1998, p. 550).	The cybersecurity unit attempts to breach the individual's systems to protect the organization's data.
Potential to harm	"The possibility of harm, rather than actual harm, since other factors may mitigate the likely harm from a betrayal (e.g., support from a sympathetic boss)" (Elangovan & Shapiro, 1998, p. 550).	The employee can be warned, penalized, or terminated for failing to properly protect organization data.
No-notice (This research)	The act of betrayal is not openly acknowledged by the betrayer (this research).	The cybersecurity unit does not disclose how or when it does vulnerability assessments to maintain efficacy and ecological validity.
Negative affect	The betrayed individual experiences negative emotions, such as anger, contempt, distrust, and fear (Grégoire & Fisher, 2008).	When finding out the cybersecurity unit is identifying a certain employee as a threat, that employee experiences negative emotions toward the cybersecurity unit.

For example, consider those voluntary tactics that attempt to identify employees who compromise the organization's confidential information (e.g., failing a simulated phishing test, plugging in a planted USB, sharing passwords, etc.). These furtive tactics could cause harm to employees (e.g., in the form of warnings, penalties, or even termination) and violate employees' core expectations (i.e., instrumental expectations of the relationship) that the cybersecurity unit will provide a safe, personal work environment; overall, this will have a negative impact.

The theory of betrayal aversion suggests that when users sense harm or the potential for harm, their relationship with an agent of protection is fundamentally altered. This suggestion is consistent with the research on marketing and organizational behavior, which suggests that harm results in a state of betrayal. When agents of protection shift to agents of harm, it has been shown that individuals feel betrayed in other contexts such as police officers, airbags, and smoke detectors (Koehler & Gershoff, 2003). In cybersecurity, if a cybersecurity unit performs vulnerability assessments that are perceived as harmful by employees, resulting in the cybersecurity unit being perceived as an agent of harm, those employees will feel betrayed (Koehler & Gershoff, 2003). Knowing whether feelings of harm alter the relationship with cybersecurity is critical because cybersecurity units are typically viewed as agents of protection—that is, trusted entities with privileged access to information about the organization, its resources, and its users.

The state of betrayal is an emotional one (Averill, 1985; Morrison & Robinson, 1997; Oatley, 1992) that results from cognitions that the individual can experience harm from the betrayer. Psychological studies of betrayal show that the state of betrayal is associated with a range of emotions that can be visceral, intense, and protracted (Browne & Finkelhor, 1986; Finkelhor & Browne, 1985; Koehler & Gershoff, 2003; Straus, 1994). These studies emphasize that a critical factor for the state of betrayal is that an individual feels the potential to be harmed (Elangovan & Shapiro, 1998). This distinction is important since an individual in such a state will feel that their well-being is being jeopardized. For example, in this context, failing a vulnerability assessment may be perceived as an impediment to workplace success. Thus, if an employee fails a phishing test, has confidential information flagged on a personal device, or is otherwise reprimanded or disciplined, they may view the cybersecurity unit as an agent of harm, leading to a state of betrayal. This would suggest that the affective response to a potential agent of harm is betrayal; hence, we hypothesize:

H2: User-perceived potential cybersecurity harm is positively related to cybersecurity betrayal.

Table 6. Active Resistance Dimensions and Behavioral Examples

Active resistance behavior	Cybersecurity example
Avoidance, defined as behaviors that evade the use of information technologies and policies of the cybersecurity function.	The individual does not report phishing attacks, does not attend or complete training, or masks documents or links in files (e.g., PDFs) in order to bypass cybersecurity screening technologies.
Abandonment, defined as behaviors that discontinue the use of information technologies and policies of the cybersecurity function.	The individual uses personal devices to complete work, subscribes to a nonwork virtual private network, or uses personal email for work.
Sabotage, defined as behaviors that deliberately undermine the use of information technologies and policies of the cybersecurity function.	The individual creates a backdoor into the system, obtains confidential records through screenshots/printouts, or disparages the cybersecurity function as incompetent.

3.3 Cybersecurity Active Resistance

The betrayal literature has suggested that employee responses to cybersecurity betrayal can range from doing nothing to circumventing the policy or seeking retribution against the betrayer (Gobin & Freyd, 2009; Grégoire & Fisher, 2008). When in a state of betrayal, people's responses can range from merely feeling disappointed to actively undermining organizational goals (Tan et al., 2021). Similarly, there is evidence in the information security and criminology literatures that even when employers follow the rules of the law and disclose that they are monitoring employees, employees may still feel threatened, leading to resistance (Lowry & Moody, 2015; Tittle, 2017). Similarly, in the context of betrayal and cybersecurity, active resistance may manifest when an employee feels betrayed, even when it is legal for the organization to carry out vulnerability assessments. This is because employees may perceive that failing or being a potential source of a cybersecurity incident may damage their standing with the organization. As a result, we operationalize active resistance to cybersecurity as manifested in three sets of behaviors: avoidance (i.e., ignoring company policy), abandonment (i.e., not using work-issued technology), and sabotage (i.e., intentionally interfering with the goals of the cybersecurity unit) (see Table 6).

The theory of betrayal aversion suggests that violations of the expectations of protection by a trusted agent of protection (e.g., a local resident or organizational member) can evoke feelings of anger, rage, denial, and avoidance (Elangovan & Shapiro, 1998; Morris & Moberg, 1994; Rachman, 2010). Even minimal feelings of betrayal can elicit fight-or-flight actions in users (Suresh et al., 2014), leading to suboptimal or counterproductive protective decisions (Koehler & Gershoff, 2003). For example, in a series of experiments, Koehler and Gershoff (2003) showed that people who felt betrayed by an agent of protection (e.g., security guard, campus police, military leader) "inflicted" higher levels of punishment on that agent than on nonprotective agents (e.g., janitor, construction worker, orchestra conductor). The betrayal literature suggests that when the actions of a cybersecurity unit elicit feelings of betrayal, even when they are taken to protect the organization (e.g., phishing as a form of employee training), employees may respond in ways that undermine cybersecurity (e.g., by intentionally violating security policies), more specifically through active resistance. Therefore, we hypothesize:

H3: Cybersecurity betrayal has a positive relationship with active resistance.

4 Research Method

To test our hypotheses, we conducted two studies: one in an organizational setting and one using panel data that was subsequently replicated in an actual organization. Study 1 used experimental vignette methods to conduct a study in a 6,000-employee organization to assess how vulnerability assessments stimulate feelings of harm and betrayal. Study 2 used a survey with panel data and data from actual employees to link cybersecurity betrayal to active resistance. We supplemented Study 1 and Study 2 with post hoc interviews.

4.1 Measurement Validation for Studies 1 and 2

4.1.1 Construct Measures.

We either contextualized existing measures or developed new measures (Hong et al., 2013; MacKenzie et al., 2011). Existing measures were modified to operationalize cybersecurity betrayal (Grégoire & Fisher, 2008; Grégoire et al., 2009). New measures were developed for cybersecurity harm and active resistance. We assessed the quality of all items in three steps: (1) we used a latent semantic analysis algorithm to analyze how semantic similarity (i.e., lexical proximity and linguistic similarity) affected the discriminant and convergent validity of the items (Gefen & Larsen, 2017); (2) after modifications, we recruited a group of academics and PhD students to perform a Q-sort with the final adjustments to the active resistance items; and (3) we conducted a pilot study with 293 participants to identify redundant items, identify opportunities for minor adjustments, and remove poorly performing items.

Table 7. Composite reliability (CR) and Interconstruct Correlations

Variable	CR	AVE	MSV	(1)	(2)	(3)	(4)
(1) Harm	0.95	0.80	0.59	**0.98**			
(2) Betrayal	0.96	0.82	0.59	0.77	**0.90**		
(3) Active resistance	0.91	0.83	0.53	0.56	0.73	**0.91**	
(4) ISec self-efficacy	0.90	0.70	0.02	-0.06	0.02	0.15	**0.95**

Note: CR = composite reliability, SD = standard deviation, MSV = maximum shared variance; Diagonal represents the square root of the AVE.

Table 8. HTMT Results

Variable	(1)	(2)	(3)	(4)	(5)	(6)
(1) Harm						
(2) Betrayal	0.81					
(3) Avoidance	0.53	0.71				
(4) Abandonment	0.53	0.70	0.82			
(5) Sabotage	0.51	0.67	0.87	0.83		
(6) ISec self-efficacy	-0.06	0.02	0.14	0.14	0.13	

4.1.2 Measurement Validation.

We used confirmatory factor analysis (CFA), EQS version 6.3, to evaluate the measurement model (Straub et al., 2004). Prior to the CFA, we verified that our data met the normality, distribution, and independence assumptions of structural equation modeling. Multicollinearity was not an issue. The highest VIF was 3.26, and the lowest tolerance level was 0.31, which is within the generally accepted guidelines that VIF should be less than or equal to 3.33 (Cenfetelli & Bassellier, 2009) and have a tolerance level greater than 0.20 (Kock & Lynn, 2012). Convergent validity was then assessed. All items had acceptable loadings (Fornell & Larcker, 1981). Composite reliability was greater than 0.70 (Nunnally, 1994), and the average variance extracted (AVE) was greater than 0.50. Next, discriminant validity was assessed. The square root of the AVEs was greater than the correlation coefficients. The item-construct loadings and cross-loadings were higher for the designated constructs than for the other constructs (in all cases, the difference was greater than 0.20) (Henseler et al., 2015). In addition, the heterotrait-monotrait (HTMT) ratios of the correlations were below the more liberal threshold of 0.90 for all constructs (Henseler et al., 2015), again indicating discriminant validity. Table 7 presents the correlations, and Table 8 presents the heterotrait-monotrait. Appendix A presents the items and cross-loadings.

CFA model fit statistics met acceptable standards (Kline, 2015). The Satorra-Bentler chi-squared value of 373.05 with 145 degrees of freedom was significant ($p \leq 0.001$). The comparative fit index (CFI) was 0.98, the root mean square error of approximation (RMSEA) was 0.06, with a 90% lower bound of 0.05 and an upper bound of 0.06, and the standard root mean square residual (SRMR) was 0.04.

4.1.3 Active Resistance as a Superordinate Construct

Active resistance was conceptualized as a superordinate construct (Polites et al., 2012) (i.e., second order and reflective) with the dimensions of avoidance, abandonment, and sabotage. We evaluated this conceptualization in three steps (Polites et al., 2012; Wright et al., 2012) (see Appendix B). First, we ran our baseline first-order factor model, with all indicators loading onto active resistance, which showed a poor fit and suggested that the construct may be multidimensional ($X^2 = 212.95$, $df = 20$, CFI = 0.96, RMSEA = 0.14, SRMR = 0.05). Then, we ran a freely correlated first-order factor model, which showed an improved fit ($X^2 = 29.26$, $df = 17$, CFI = 0.99, RMSEA = 0.04, SRMR = 0.05), as well as a significant indicator of loadings, suggesting multidimensionality and convergent validity. Next, we compared the unconstrained models to constrained models between sets of factors, which showed a significant X^2 change and supporting discriminant validity. Finally, we ran our parallel model ($X^2 = 29.73$, $df = 17$, CFI = 0.99, RMSEA = 0.04, SRMR = 0.04), Tau equivalent model ($X^2 = 27.81$, $df = 15$, CFI = 0.99, RMSEA = 0.04, SRMR = 0.04), and congeneric model ($X^2 = 20.89$, $df = 13$, CFI = 0.99, RMSEA = 0.04, SRMR = 0.02), with all three showing an acceptable fit. Collectively, these analyses indicate that active resistance was appropriately modeled as a superordinate construct.

5 Study 1: Vulnerability Assessments as a Source of Harm

To examine whether vulnerability assessments cause harm (Proposition and Hypothesis 1), we used a vignette field experiment and manipulated the type of vulnerability assessment (i.e., email scanning, phishing, and personal device scanning). The selection of

vulnerability assessment techniques was informed by a panel of seven chief information security officers (CISOs) from Fortune 500 companies. Our experts identified common and potentially harmful tactics among several choices identified as vulnerability assessment techniques from the Penetration Testing Execution Set standards (PTES). Among the PTES choices, the tactics were explicitly selected for this study because they mapped to (Appendix C, Table C2) and were noted as being used in vulnerability assessments by four security researchers and our panel of CISOs. Of the choices, the CISOs identified email scanning, phishing, and personal computing device scanning as the most widely used in industry (Appendix C). The CISOs suggested that email scanning was the least harmful, that phishing could be perceived as moderately harmful, and that personal device scanning would most likely be perceived as harmful.

5.1 Experimental Vignette Development

Consistent with the related work (Jasso, 2006), we used an "actual derived cases" approach to develop vignettes that elicit different feelings of harm in participants (Aguinis & Bradley, 2014). After developing our vignettes, we returned to our expert panel of CISOs and asked them to assess their realism and generalizability in an organizational setting (Appendix C). Following a review by our expert panel, we worked directly with a cybersecurity compliance office at a real organization to gather feedback from 228 employees on our initial vignettes. After reviewing the vignettes, employees at the organization rated the level of perceived harm (M = 4.93, SD = 1.65, Likert scale *strongly disagree* = 1, *strongly agree* = 7) and rated the realism of the vignettes as 8.2 on a 10-point sliding scale (i.e., 0 = *not realistic at all* to 10 = *completely realistic*). The results from our CISO panel and pretest ratings indicate that the vignettes are ecologically valid and were perceived as harmful by cybersecurity professionals and employees who would be subject to vulnerability assessments in an actual organization.

5.2 Experimental Procedures

We used our vignettes to construct manipulations in coordination with an organization that used vulnerability assessments. This organization was actively conducting phishing simulations and scanning and exfiltrating information with payloads from all connected devices while also using various anomaly detection systems to monitor and assess the behavior of more than 6,000 employees in the southeastern United States. We manipulated the type of vulnerability assessment tactics used on actual employees who were subject to vulnerability assessments in the organization.

The experimental vignette, which was hosted on Qualtrics, was administered in six steps: (1) We recruited organizational participants via email in waves (500 email requests for the field experiment and post hoc interviews) and incentivized participants with the chance to win a virtual reality gaming headset or a $100 gift card. (2) We measured whether the participants expected protection from the organization's cybersecurity unit (M = 5.78, SD = 0.94, Likert scale 1 = *strongly disagree* to 7 = *strongly agree*) before reading the vignette. (3) Participants were randomly assigned to one of the treatment groups (i.e., control, low, medium, or high). (4) Participants were presented with the vignette. (5) We verified that the vignette was perceived as a real possibility. (6) We instructed participants to imagine themselves as the fictional character while responding to the survey instrument measuring harm on a seven-point Likert scale (1 = *strongly agree* to 7 = *strongly disagree*).

The vignettes defined the vulnerability assessment tactic for the participant and explained how the cybersecurity unit used the tactic: (1) Email and internet scanning is the act of cybersecurity tracking, observing, and identifying vulnerabilities in the employee computing activities on organization-owned computing devices. (2) Phishing is the act of cybersecurity carefully crafting emails that are sent to selected employees with the goal of convincing them to disclose network credentials and/or download a malicious file. (3) Personal device scanning is an act of cybersecurity that tracks, observes, and identifies vulnerabilities in an employee's computing activities on their own personal computing devices (e.g., smartphone, tablet, smartwatch, etc.) connected to the organization's network.

Three one-item manipulation checks were used to ensure that the participants understood the vignette and manipulation (Marett, 2015). The participants were asked to rate the following statements as "true" or "false": (1) "Jamie's confidential information was compromised by the cybersecurity unit …"; (2) "According to the vignette a … compromised Jamie's confidential information"; and (3) "Jamie's confidential information was compromised by a …" The participants were given the choice regarding whether the cybersecurity unit compromised confidential information through the assessment tactic. The choices were none, email scanning, phishing, or personal device scanning. If a participant did not answer correctly, the manipulation check was noted as failed, and the participant was removed from the study. To assess whether participants read the vignette, we measured the time spent reading the treatment. The mean time spent reading the vignette was 26.1 seconds, which is sufficient to read an eight-line vignette.

5.3 Experiment Participants

In total, 297 participants completed the experiment. Forty-nine participants failed either the manipulation check or the realism check (i.e., rating on a scale of 0 = *not realistic at all* to 10 = *completely realistic*). Of the remaining 248 usable responses, approximately 60.9% were female; the average participant was 41.9 years old, had an average of 23.0 years of computer experience, had been employed by the current organization for 10.6 years, and had experienced an average of 2.9 cyberattacks.

5.4 Experiment Results

Using SPSS 26.0, we performed an ANCOVA to analyze perceived cybersecurity harm as the dependent variable (Table 9). We found that vignettes in which the user was subjected to a vulnerability assessment stimulated feelings of harm across all three conditions while controlling for the time at the current organization, past cyberattack victimization, and self-efficacy in information security compared with the control group. The empirical evidence supports Hypothesis 1 ($F_{(3, 248)} = 21.72$, $p \leq 0.001$) because the control condition mean is lower than the three treatment means (control condition, $M = 3.05$, $SD = 1.39$, $N = 57$; phishing condition, $M = 5.77$, SD = 1.34, $N = 55$; scanning personal devices, $M = 5.42$, SD = 1.43, $N = 65$; email scanning condition, $M = 5.06$, SD = 1.55, $N = 71$), demonstrating that no-notice vulnerability assessments stimulate feelings of harm by the cybersecurity unit. Post hoc comparisons indicated that all forms of vulnerability assessments—email scanning ($p \leq 0.001$, Cohen's $d = 1.37$), phishing ($p \leq 0.001$, Cohen's $d = 1.99$), and scanning personal devices ($p \leq 0.001$, Cohen's $d = 1.37$)—resulted in perceived cybersecurity harm, with large effect sizes compared with the control vignette. Additionally, phishing as a tactic differed significantly from email scanning ($p = 0.012$). Cohen's d was 0.49, suggesting a moderate effect size for that comparison.

5.5 Study 1 Post Hoc Experiment Interviews

To gain insight into why vulnerability assessments may have created feelings of harm, we sought input via interviews from another wave of 500 randomly selected employees across all departments and levels (i.e., executives, managers, workers, etc.) of our partner organization. We received responses from 367 employees (73.4%) and asked them how they felt about their organization's use of email scanning, phishing, and personal device scanning for vulnerabilities and training. Table 10 shows representative quotes.

Table 9. Results of ANCOVA of Vulnerability Assessment Tactics

Predictor	Sum of squares	*df*	Mean square	*F*	*p*	Partial η^2
Intercept	271.67	1	271.67	137.26	< 0.001	0.36
Assessment tactic	65.16	3	21.72	10.97	< 0.001	0.12
Time at current organization	2.02	1	2.02	1.02	0.314	0.00
Past cyberattack victimization	13.35	1	13.35	6.75	0.010	0.03
Self-efficacy in information security	0.50	1	0.50	0.25	0.616	0.00
Error	477.01	241	1.98			
Adjusted $R^2 = 0.36$						
Post Hoc Analysis					*p*	Cohen's *d*
Control condition	Email scanning				< 0.001	1.37
	Phishing				< 0.001	1.99
	Scanning personal devices				< 0.001	1.37
Email scanning	Control condition				< 0.001	1.37
	Phishing				0.012	0.49
	Scanning personal devices				0.199	0.24
Phishing	Control condition				< 0.001	1.99
	Email scanning				0.012	0.49
	Scanning personal devices				0.205	0.25
Scanning personal devices	Control condition				< 0.001	1.37
	Email scanning				0.199	0.24
	Phishing				0.205	0.25

Table 10. Post Hoc Interviews on Vulnerability Assessments as a Source of Harm

Vulnerability assessment as a source of harm	Exemplar quotes of cybersecurity as a source of harm
Email scanning	"It's almost as if I opened my door to a sociopath." "The duration of the monitoring is what bothers me—if the goal was to see if security could be penetrated—once it was, the hole needs to be plugged—not continue to be exploited … [T]hat it was ongoing is what bothers me." "Confidential information, which if reviewed and kept, or searched for more specifically, I would feel that the cybersecurity team acted inappropriately." "I think the cybersecurity department is falsely proclaiming their duties." "Employees need protection from and strict regulations on cybersecurity because of the confidential information they have access to."
Phishing attack	"It is not unheard of for cybersecurity and contracted consultants to abuse their station to gain sensitive information." "It seems more like a personal attack to see if he gives out confidential information." "Jamie needs to process, vent, and not go back to a created toxic environment."
Scanning personal devices	"The vignette is [a] real possibility and scary to me." "If it was a test, I don't think it should have been done. It could have been tested on something less important or even a simulation with fake information." "Monitoring the content/activity described feels like a violation of trust. It doesn't instill confidence that the organization trusts its employees but is seeking to undermine or police their activities."

Our participants shared that no-notice vulnerability assessments made them feel vulnerable to cybersecurity, saying that "cybersecurity can get you," "they can cross any line," and "it calls into question what they are doing." They also expressed that they would "feel threatened or not trust cybersecurity" and would respond to their tactics by "trying to undermine or control [cybersecurity's] activities." Given the support provided by our experiment and post hoc interviews, we moved on to Study 2, which examines the consequences of cybersecurity betrayal.

6 Study 2: The Consequences of Cybersecurity Harm and Betrayal

Study 2 examined whether perceived harm induced by vulnerability assessments has a positive relationship with betrayal. We collected panel data from full-time employees of the organization. We then randomly assigned the participants to view our vignettes and used seven-point Likert scales (1 = *strongly disagree* to 7 = *strongly agree*) to collect perceptions of cybersecurity harm, betrayal, active resistance, and the control variable of information security self-efficacy (see Appendix B). As noted above, these measures demonstrated adequate discriminant and convergent validity.

6.1 Participants

Our participants were drawn from a panel provided by Qualtrics. Eligibility requirements included (1) being a native English speaker, (2) working full time, (3) working in the United States, and (4) using a computer for work. We also required the participants to confirm that the vignette was a real possibility in their organization. Of the 615 participants who passed our qualifying questions and realism check, we removed 119 unusable responses: 20 participants did not complete the survey, 2 failed the attention check, 46 failed the manipulation check, 47 failed the realism check, and 4 spent too little time on the questions (< 4 minutes) (DeSimone et al., 2015). Of the 496 usable responses, approximately 62.0% of participants were female, the mean age of participants was 42.5 years, they had a mean of 21.5 years of computer experience, they had been employed at their current organization for 10.0 years, and they had experienced a mean of 2.6 cyberattacks. The average time spent reading the vignette was 26.1 seconds, indicating sufficient time to read the eight-line vignette.

In this manipulation check, participants were asked on a nominal scale (0 = *penetration test*, 1 = *hack*, and 2 = *does not matter*) whether their perception of the vignette was a penetration test, here looking for differences in their perceptions of harm (M = 5.33, SD = 1.50); a hack (*M* = 4.90, *SD* = 1.85); or did not matter (*M* = 4.89, *SD* = 1.72).

We found no significant differences between the groups (F (2, 496) = 1.804, p = 0.166). This manipulation control ensured that the study participants understood that the cybersecurity unit was conducting the vulnerability assessment for training and that the unit was a potential source of harm.

6.2 Analysis of the Structural Model

We used AMOS version 6.3 to test the hypotheses (Figure 4). The results are presented in Table 9. The Satorra-Bentler chi-squared value of 810.93 with 260 degrees of freedom was significant ($p \leq 0.001$). The CFI was 0.96, the RMSEA was 0.06, with a 90% lower bound of 0.06 and an upper bound of 0.07, and the SRMR was 0.08. All fit statistics indicated that the structural model met the acceptable criteria (Kline, 2015). We also evaluated the presence of common method variance using the common latent method factor approach and found that it did not significantly affect our results (Williams et al., 2010) (see Appendix D).

6.3 The Effects of Cybersecurity Harm on Betrayal

Our analysis supported H2 (ß = 0.76; t = 19.83; $p \leq$ 0.001) and H3 (ß = 0.67; t = 15.61; $p \leq$ 0.001), indicating that when employees perceive cybersecurity harm, they feel betrayed by the cybersecurity unit and are more likely to engage in active resistance.

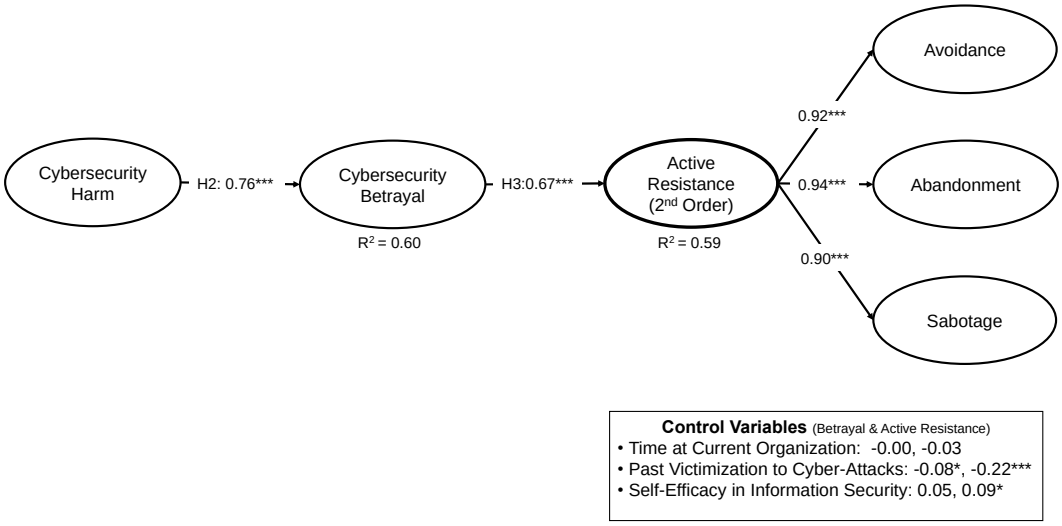
6.4 Replication of Study 2 in Organization

To verify the robustness of the findings and add ecological validity, the study was replicated in a real organization. We surveyed a group of employees in the organization and asked them how they would feel if they were subjected to vulnerability assessments conducted via email scanning, phishing, or the scanning of personal devices. The results from our model replicating our research model were consistent with the results from the panel data, as shown in Figure 5.

6.5 Study 2 Post Hoc Interviews

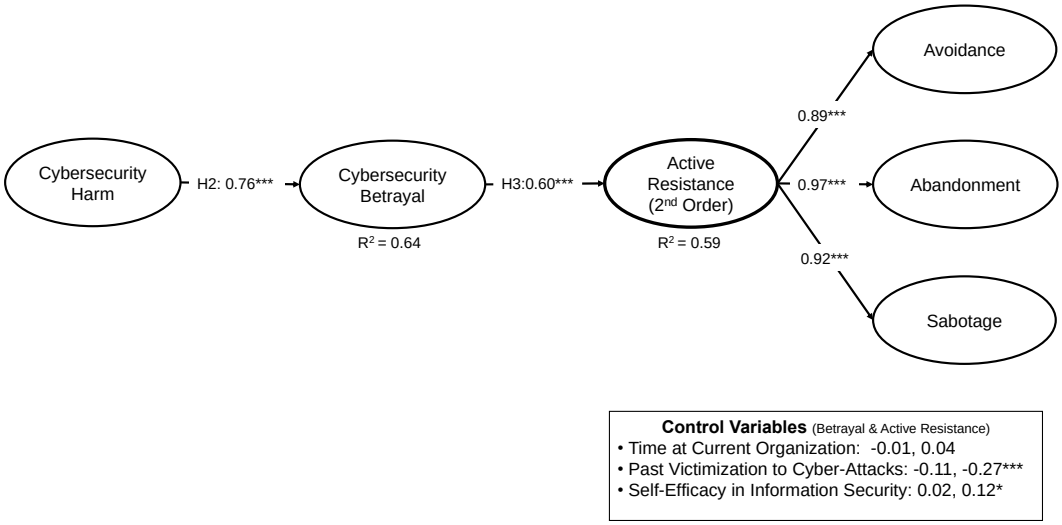
To gain additional insights, we asked our participants post hoc about how cybersecurity betrayal shapes responses to no-notice vulnerability assessments (Table 11).

Our post hoc interviews revealed that even when participants knew that a vulnerability assessment was done for protection, they still felt betrayed. "Technically, it's fair and legal …" one participant noted, who went on to point out, "making their access to all of his/her documents/info seem like a betrayal." Another participant felt that cybersecurity violated pivotal expectations by falsely proclaiming their duties and argued that they had "all the reason to no longer trust this cybersecurity team." More than one participant offered a global negative assessment of this type of assessment. One proclaimed: "These tests are set up with ill intent." Another stated that the tactic "created a toxic environment." A third stated that they would feel "upset if my private devices were monitored." Others reported that the vulnerability assessment made them feel "embarrassed and attacked" and that the user is "SOL if something happens, which isn't fair, but is corporate reality." Overall, our post hoc interviews confirmed that the participants believed that employees could feel betrayed by the cybersecurity unit in an organizational context.



Note: ***p < 0.001, **p <0.01 *p < 0.05; Model Fit: X2 = 810.93, df = 260, CFI = 0.96, RMSEA = 0.06, SRMR = 0.08

Figure 4. Study 2: Research Model Testing Results



Note: ***p < 0.001, **p <0.01, *p < 0.05; Model Fit: X2 = 565.17, df = 260, CFI = 0.93, RMSEA = 0.06, SRMR = 0.08

Figure 5. Study 2: Replication in Actual Organization Results

Table 11. Post Hoc Interviews of Theoretical Relationships

Relationship	Exemplar quotes
Vulnerability assessment and betrayal	"Technically, it's fair and legal; I just doubt an employee would read the whole contract and the 'fine print' making their access to all of his/her documents/info seem like a betrayal." "I believe the team should have your trust until given a reason not to have it. Jamie has all the reason to no longer trust this cybersecurity team." "I believe that once cybersecurity told him they do this to protect his information better, he would be less embarrassed and less attacked, but in the real world, they don't." "A common and still unfair vignette in work today." "Every single move Jamie makes no matter how slight the company knows … [T]he company has her [by] the short hair, whether she likes it or not. It's not fair but corporate reality." "This seems really out of line by cybersecurity." "These tests are set up with ill intent even if just testing or necessary for the information protection." "This spoke of possible misuse of power by the cybersecurity department; the more I doubted the use of the perimeter test on people." "Security needs to be a step ahead of attackers, but at what cost!" "I believe as long as the employee is transparent and performs as instructed, they have nothing to hide, but cybersecurity should realize the same and let them know how they are being scrutinized by the cybersecurity division."
Active resistance	"Jamie should take steps fast to protect her files and confidential information from penetration testing." "Employees need protection from and strict regulations on cybersecurity because of the confidential information they have access to." "Do not trust the cybersecurity they can get you." "We are SOL if something happens, which isn't fair to the employee—either buy me a secure device or tell me not to do work on my phone." "There are other ways to monitor an employee if there are suspicions without breaching their trust."

The post hoc interviews also show that the participants were willing to take actions to protect themselves from the cybersecurity unit. One emphasized that violations of protective expectations meant that they had to "be 'on guard' at all times when communicating." Others expressed the opinion that "cybersecurity needs checking and monitoring" and that an employee should have the right to "… speak to his higher-ups" about being monitored. Because one participant no longer felt "trust in cybersecurity," he wanted the organization to provide "a secure device or tell me not to do work on my phone." Overall, our participants supported our view that betrayed employees may take action against the cybersecurity unit.

6.6 Post Hoc Tests of Studies

Mediation analyses: Since mediation is present in our model, we tested the mediating effect of harm (i.e., mediator) on the relationship between vulnerability assessments (i.e., independent variable) and betrayal (i.e., dependent variable). To test this, we performed a post hoc ANCOVA with betrayal as the dependent variable and a mediation test using PROCESS v 3.5 in SPSS v. 26.0 to understand the mediating role of cybersecurity harm. The results of the ANCOVA for the treatment ($F_{(3, 248)}$ =3.69, p < 0.027) were significant. Post hoc pairwise comparisons revealed that the only significant effect was between email scanning and phishing (p = 0.008). PROCESS Model 4 was used to test the mediating effect of harm between vulnerability assessment tactics and cybersecurity betrayal. Note that the type of action of the vulnerability assessment was entered as a multicategorical variable (i.e., dummy coded), which PROCESS can handle. The results of the path comparing email scanning (reference group) to phishing (a_1 path) were significant (t = -2.40, p = 0.017), and comparing the email to personal devices (a_2 path) was not significant (t = -1.11, p = 0.267), as was the result of our experiment analyzing the mediating effect of harm. The result of the path from cybersecurity harm to betrayal (b path) was significant (t = 13.82, $p \leq 0.001$). The result did not show significant direct paths (c_1 and c_2) from treatments to betrayal (t = -1.32, p = 0.19 and t = -1.04, p = 0.30). The bootstrapping results showed that the mediated path (a_1b) for comparing email scanning to phishing supported indirect-only mediation (point estimate = -0.49, 95% CI = -0.88 to -0.11) (Zhao et al., 2010). We also tested betrayal as a mediating variable between harm and active resistance. We used AMOS to perform a mediation analysis using bootstrapping. The results showed that the mediated path (ab) from harm to active resistance was significant ($p \leq 0.001$, 95% CI = 0.37 to 0.62), but the direct path (c) was not significant (p = 0.783, 95% CI -0.12 to 0.14). Our results showed that harm is an important mediating variable between treatment and betrayal, and betrayal is an important mediating variable between harm and active resistance.

Several subjects failed the manipulation check. To test whether the results were robust, an ANCOVA was analyzed with the subjects who were (previously) removed. The results of the experiment, including all subjects (N = 298), were consistent with the results without the manipulation check. The empirical evidence supported H1 ($F_{(3, 298)}$ = 22.68, $p \leq 0.001$), and post hoc comparisons showed that all forms of vulnerability assessments—email scanning ($p \leq 0.001$), phishing ($p \leq 0.001$), and personal device scanning ($p \leq 0.001$)—were significant compared with the control group. In addition, phishing as a tactic was significantly different from email scanning (p = 0.004). These additional tests demonstrate the robustness of our results and support our theoretical underpinnings.

Finally, we ran structural models from Study 2 with a direct effect from harm to active resistance. The model with the direct path from harm to active resistance was consistent with this additional path in both pools of participants (Qualtrics and the organization). Furthermore, the direct path from harm to active resistance was not significant (ß = 0.02, t = 0.20, p = 0.845) in the structural model from Study 2. Again, we empirically confirmed the theoretical relationships in our research model.[1]

Mitigating unintended consequences: We conducted a series of post hoc moderation analyses on our model to probe theoretically relevant variables that might reduce the negative consequences of harm. We selected the variables of trust, information security mindfulness, and information security self-efficacy as potential moderators. We proposed that these variables would dampen the relationships between vulnerability assessments, harm, betrayal, and active resistance. We then conducted our moderation analysis by including these variables in the research model.

Trust, information security mindfulness, and information self-efficacy were theorized to dampen the relationships in our moderation analysis. Trust was operationalized as three different types of trust: trust in people (i.e., the cybersecurity function), trust in technology (i.e., cybersecurity technology safeguards), and trust in the organization (i.e., situational normality and structural assurances). Trust has been shown to reduce uncertainty in situations that make individuals vulnerable (Pavlou et al., 2007; Sollner et al., 2016). Similar to other contexts such as e-commerce, we propose that individuals who trust the cybersecurity function will understand why vulnerability assessments are necessary for the organization. Information security mindfulness is a state

[1] We thank the senior editor and reviewers for the suggested post hoc analyses.

in which individuals are in a more active and alert state that allows them to notice multiple perspectives (Langer, 1989). In the context of information security, mindfulness is also shown to encourage employees to pause and understand the situation (Jensen et al., 2017) as well as engage in proactive behaviors (Burns et al., 2019). Therefore, individuals who are interested in information security may take a step back and understand why the cybersecurity department is conducting vulnerability assessments. Self-efficacy in information security (SEIS) refers to the belief in "one's ability to protect information and information systems from unauthorized disclosure, modification, loss, destruction, and lack of availability" (Rhee et al. 2019, p. 817). SEIS has been shown to influence cybersecurity behaviors (Rhee et al., 2009). The strength of these beliefs has been shown to be an important factor influencing an individual's cybersecurity compliance behavior and response to threats (Bulgurcu et al., 2010; Rhee et al., 2009; Yoo et al., 2020). Individuals with higher SEIS may feel more confident that the cybersecurity unit is protecting them by conducting vulnerability assessments and be more likely to understand the purpose.

Including our variables of interest as moderators resulted in a moderated mediated model (Appendix E). To test this model, we used PROCESS in R version 4.3.1 as it is well suited for testing these types of models. Specifically, we isolated each type of vulnerability assessment: scanning email, phishing, and scanning personal devices. We used Model 59 in PROCESS, which tested the moderated mediated relationship between vulnerability assessments, harm, and betrayal. Model 59 in PROCESS tests the moderating effect of our variable of interest on all paths (a', b', and c'). The results show significant negative moderation for trust in people (email scanning: *n.s.*; phishing: *n.s.*; scanning personal devices: ß = -0.11, t = -3.38, p = 0.001), trust in technology (email scanning: *n.s.*; phishing: ß = -0.10, t = -1.99, p = 0.049; scanning personal devices: *n.s.*), and trust in the organization (email scanning: *n.s.*; phishing: *n.s.*; scanning personal devices: ß = -0.09, t = -3.02, p = 0.003) on the path between harm and betrayal only (i.e., b'), thus dampening the relationship between harm and betrayal for certain vulnerability assessments. The results for mindfulness only showed a significant positive moderating relationship for the path (i.e., a') between scanning personal devices and harm (ß = -0.11, t = 2.57, p = 0.011), thus amplifying the relationship between scanning personal devices and harm. The results for information security self-efficacy showed a negative moderating effect for email scanning (ß = -0.10, t = 2.12, p = 0.036) only. No other paths showed significant moderation.

We also conducted an additional post hoc test using Model 59 in Process on the variables of interest for moderating the path to harm, betrayal, and active resistance. The results showed a significant negative moderating effect for trust in people (ß = -0.07, t = -3.97, $p \leq 0.001$), trust in technology (ß = -0.07, t = -3.61, $p < 0.003$), and trust in the organization (ß = -0.05, t = -3.04, $p < 0.003$) on the path between harm and active resistance (i.e., c'). Thus, all forms of trust dampened the relationship between harm and active resistance. The results also showed a significant negative moderating relationship for information security mindfulness on the path between betrayal and active resistance (ß = -0.07, t = -2.07, $p < 0.039$). This dampens the relationship between betrayal and active resistance. Interestingly, there were no significant moderating effects for cybersecurity self-efficacy on any paths in the model. All other moderations were not significant in the post hoc analysis.

7 Discussion

We were interested in understanding the impact of vulnerability assessments on employees' feelings toward the cybersecurity unit and their subsequent cybersecurity behavior. We conducted a pair of studies using mixed methods to investigate vulnerability assessments, cybersecurity harm, betrayal, and the unintended consequences of these constructs. We showed that employees view vulnerability assessments as harmful compared with other forms of protection provided by cybersecurity units. Our post hoc interviews further revealed that even when employees understand the purpose of the assessment, they may still feel harmed. Study 2 showed that when employees feel harmed, they feel betrayed and engage in active resistance, confirming our theoretical causal chain of unintended consequences. In addition, the post hoc tests demonstrated the robustness of our theory and model. Overall, the analysis suggests that users feel harmed and betrayed and engage in actions that undermine the cybersecurity units of organizations.

To the best of our knowledge, our research is the first to shed light on the negative consequences of performing vulnerability assessments on employees and their data. Rooted in the theory of betrayal aversion, our findings suggest that regardless of the tactic (i.e., email scanning, phishing, or personal device scanning), these tactics violate employees' expectations and create feelings of harm by the cybersecurity unit. Based on the theory of betrayal aversion, employees reevaluate their relationship with cybersecurity, realizing that an agent previously viewed as protective may be an agent of harm. Consistent with our expectations, we found that assessments that rely on psychological manipulation, such as phishing (Wright et al., 2014), are perceived as causing more harm. Because we found that vulnerability assessments jeopardize the relationship between the cybersecurity unit and the organization's employees, we suggest that future cybersecurity work should focus on preserving, building, or repairing the relationship between the cybersecurity unit and employees.

When organizations perform vulnerability assessments of employees, they should be aware of the potential positive and negative consequences. On the one hand, regulators

recommend vulnerability assessments, and organizations have a legal right to conduct cybersecurity vulnerability assessments when disclosed in policy. On the other hand, our work implies that cybersecurity units should anticipate that such assessments can yield consequences that may be detrimental to their objectives of securing the organization. Thus, organizations must consider whether the benefits of conducting vulnerability assessments outweigh the costs. Often, cybersecurity assessments such as those recommended by the NIST (e.g., NIST 800-115: Technical Guide to Information Security Testing and Assessment) focus on identifying perceived threats in the organization; these frameworks might consider incorporating how such vulnerability assessments align with the organization's mission and employee well-being.

Organizations need to ask themselves if conducting a mock phishing test and monitoring employee activities are worth a minimal increase in protection when there are other ways to conduct vulnerability assessment more prudently in a "safe" environment. The organization in the present study has changed its approach toward creating, designing, and conducting phishing tests, such as stopping them during the COVID-19 pandemic and following up with in-person training sessions for entire workgroups when an employee fails. The company also warns employees that a test will be conducted sometime in the next quarter, which maintains some ecological validity but does not undermine employees' expectations of protection. Also, for personal devices, the organization began informing users in training sessions that when they connect their devices, data running through the network would be visible. They also offered free authentication tokens (e.g., devices other than personal phone) for employees who were uncomfortable connecting their devices and held training sessions to help employees configure their own networks at home—all designed to increase trust and allow employees to learn about the cybersecurity unit and why it was testing people and systems.

Finally, our post hoc moderation analyses show that there are important conditional factors when conducting vulnerability assessments. While vulnerability assessments are necessary, if trust can be established through initiatives that promote the cybersecurity department, its technologies, and the support of the organization, it may diminish the negative consequences of vulnerability assessments. Additionally, the results show the bright and dark sides of mindfulness. On the one hand, users may be mindful of how their devices interact with vulnerability assessments, amplifying feelings of harm when personal devices are scanned. On the other hand, more mindful users may pause and understand the ramifications of engaging in active resistance. This underscores the importance of the relationship between the cybersecurity unit and employees in establishing trust and different approaches to training employees to increase their mindfulness and self-efficacy.

7.1 Theoretical Implications

Our findings have important theoretical implications. First, we provide insights into employees' negative reactions to cybersecurity units conducting vulnerability assessments on the human element of the organization. Although cybersecurity research offers a rich understanding of how certain mechanisms (e.g., fear, motivation [Boss et al., 2015; Johnston et al., 2015], accountability [Vance et al., 2015], and deterrence [D'Arcy et al., 2009]) evoke compliance, we offer a counter-narrative of how such mechanisms unintentionally evoke active resistance behaviors that undermine compliance. By doing so, the present study highlights the importance of the emerging dark side of the cybersecurity literature, which has suggested that organizational actions such as greenwashing or stress-inducing cybersecurity tactics (D'Arcy et al., 2014) can undermine the goal of protecting corporate data. Such work is important because research that links unintended consequences to weakened firm cybersecurity can help identify ways to create safer and more trustworthy cybersecurity practices, as our study did at the focal organization.

Second, by introducing cybersecurity betrayal into the literature, we draw attention to the negative consequences of cybersecurity, showing how it can be perceived as an agent of harm rather than an agent of protection. Existing cybersecurity research has focused on breaches and noncompliance as a byproduct of either the failure of users to adopt security awareness and training programs and policies (D'Arcy et al., 2014) or the success of cybercriminals (Wright et al., 2014). In contrast, our research draws attention to the importance of the perceptions of the internal cybersecurity unit. Our focus on employee reactions to cybersecurity in an organization shows that when employees become aware of different tactics to secure the organization, they often feel betrayed because they no longer view the organization's protective expectations and rules as benign. Although the current research has focused on perceived vulnerability assessments by the cybersecurity unit that stimulate harm, betrayal, and active resistance as a form of noncompliance, there may well be unknown positive effects of perceiving cybersecurity as a protective agent, such as the more frequent reporting of threats. Future work should examine whether the perceived positive and negative consequences of cybersecurity policies, assessments, and technologies further diminish or enhance the protective relationship between the cybersecurity unit and employees.

Third, beyond cybersecurity, our work contributes to IS resistance research by going beyond examining the disruption caused by the introduction or implementation of a new system in an organization (Craig et al., 2019; Rivard & Lapointe, 2012). Rather, we show that the way we use established technologies, especially those that create potential harm and betrayal for employees, can create resistance to established technologies (Burleson et

al., 2021). Our post hoc interviews demonstrate the need to examine how the repurposing or evolution of existing information systems can lead to unintended consequences, such as active resistance. This finding is particularly relevant for more advanced technologies, such as artificial intelligence, which can evolve without human intervention (e.g., human-programmed updates, patches, etc.) (Rai et al., 2019). A broader understanding in IS research of how to effectively deploy these technologies is needed as they become more widely adopted in organizations.

Finally, the present research contributes to the literature on betrayal. Numerous studies have focused on betrayal because of aberrations (Grégoire & Fisher, 2008; Reina & Reina, 2006). Our findings illustrate how a routine set of actions can lead to betrayal and negative outcomes for organizational cybersecurity. For example, we have found that following betrayal, employees respond with more active forms of resistance, namely avoidance (i.e., circumventing the organization's network security), abandonment (i.e., not using work-issued technology), and sabotage (i.e., intentionally disrupting the goals of the cybersecurity department). Thus, when organizations routinely fail to meet core protective expectations, employees often respond by proactively attempting to undermine organizational protocols. It would be interesting to examine, for example, if the outcomes of routine violations of pivotal expectations be different for cybersecurity and human resources. Future research evaluating whether such responses result from the vector of betrayal may reveal that different techniques work better in different departments.

7.2 Practical Implications

Our work supports the growing movement among cybersecurity practitioners to consider the impact of cybersecurity tactics on employees and their responses (e.g., human-centric cybersecurity; McKee, 2021). For example, in 2020, one of the largest cybersecurity practitioner conferences, RSA, focused on this human-centered approach to security and explored how human-centric cybersecurity can inform organizations' cybersecurity tactics. This movement was spurred by conversations suggesting that cybersecurity professionals do not fully understand the impact of cybersecurity interventions on employees and their desire to be good stewards of the people, data, and information in their organizations (McKee, 2021).

Our research provides a rigorous evaluation of cybersecurity harm that puts the emerging human-centric cybersecurity movement into practice. To this end, we provide cybersecurity professionals with evidence that their concerns about the impact of vulnerability assessments on the relationship between the cybersecurity unit and employees are well founded. We emphasize to CISOs and managers that there is a trade-off between using these tactics to secure an organization and provoking active user resistance, which can then threaten cybersecurity itself. Because social engineering tactics are based on manipulating individuals and exploiting trust (Wright et al., 2014), practitioners need to be aware that such tactics can cause perceived harm to individual employees and create threats to the cybersecurity ecosystem.

Second, our work suggests that practitioners should be careful about how they use vulnerability assessment tactics. Although some industry conversations focus on how to mimic cybercriminal techniques (Bacudio et al., 2011), we believe that cybersecurity professionals need to work with employees to consider how to use such techniques without evoking feelings of betrayal. One of our post hoc participants pointed out a way to do this: provide opportunities for employees to voice concerns about the use of vulnerability assessment techniques. By engaging employees in processes such as reviewing vulnerability assessment tactics, providing comprehensive internal documentation, debriefing employees more effectively, and sharing insights about what they have learned through the use of vulnerability assessment tactics, cybersecurity could mitigate the impact of betrayal on the relationship between cybersecurity and employees.

Finally, practitioners should explore technical and policy solutions that help avoid feelings of harm and betrayal by the cybersecurity unit. For example, organizations can offer alternatives to connecting personal devices for multifactor authentication. A simple hardware token—a small physical device that looks like a key fob or credit card—can authenticate a user to a system without requiring the person to connect personal devices. More comprehensive solutions that protect employee anonymity, such as homomorphic encryption (i.e., conducting analysis on unencrypted data), provide a new way for cybersecurity professionals to gain the trust of employees in organizations.

7.3 Limitations

Before discussing future research directions, we consider the limitations of our work. One limitation may be response bias. To address this issue, we designed our studies to reduce response bias, such as ensuring participant anonymity and varying the vignettes participants view (Aguinis & Bradley, 2014; Jasso, 2006). Second, we examined only a few of the most common vulnerability assessment tactics used in organizations. Although our selection of tactics may limit the generalizability of our findings, our choice of tactics was based on the opinions of an expert panel of CISOs. Third, our work used the experimental vignette method, which allowed us to manipulate users' vulnerability assessment tactics. This method is often used when studying topics that could be detrimental to an organization, such as revealing the organization's use of vulnerability assessment. To overcome this limitation, we conducted a series of pre- and post-study analyses. For example, we

developed our scenarios with input from a panel of CISOs and actual users in an organization. We measured perceptions of the scenario (i.e., perimeter test, hack, neither), whether subjects believed this could happen, and whether users felt protected in their organization before administering our experimental vignette. In future work, researchers may need to assess whether less commonly used tactics elicit different responses and partner with organizations for field experiments.

7.4 Future Research Directions

Our work provides exciting avenues for future research on betrayal and social exchange. Research examining the group effects of betrayal could shed light on whether an actual act—or the mere possibility of betrayal—leads to active resistance to the cybersecurity unit among employees. For example, if a colleague is fired from an organization for failing to identify phishing emails, will other employees begin to engage in active resistance behaviors out of resentment? We suspect that this outcome is likely given the power of negative emotions/effects on behavior (Baumeister et al., 2001). Similarly, could feelings of betrayal have a contagion effect, such that acts of betrayal lead to widespread dissatisfaction that ripples through the organization (Barsade, 2002), stimulating more counterproductive work behaviors (Kelly & Barsade, 2001) that harm the organization and, hence, leading to higher turnover rates?

In addition, researchers need to examine whether an emotional spillover of cybersecurity betrayal affects the relationship between the organization as a whole and employees (e.g., job satisfaction). For example, does the impact of cybersecurity betrayal affect different facets of employees' lives beyond their work? For instance, an individual might perceive that cybersecurity crosses from the work domain into the personal domain, with little recourse for the employee to stop these actions. Therefore, this perception could lead to increased feelings of work-life conflict that affect relationships outside of work. Additional moderators, such as positive and negative emotional moderators, coping responses, work attitudes, and other interjections via moderation of our initial theory, should prove fruitful for future cybersecurity and IS research.

More work is needed on how cybersecurity interventions can have unintended consequences for organizational cybersecurity and the organization itself. Although studies have examined how technologies designed to protect employees can sometimes fail (Vance et al., 2018), little research has examined how technologies that fail can stimulate feelings of harm in employees. Future IS research could clarify how the implementation or performance of an information system stimulates feelings of betrayal and resistance, as well as feelings of protection and poor performance (Craig et al., 2019). Also, although we did not set out to create a taxonomy of vulnerability assessment tactics and employee responses, future researchers could use our vulnerability assessment grid to further explore the differential effects of these tactics on employees. By acknowledging that the features of systems can elicit multiple responses, research will provide a richer view of the effects of technology in sociotechnical systems.

Finally, our work suggests the need to examine how organizations should consider human-centered cybersecurity policies. Vulnerability assessment tactics are an essential means of identifying and mitigating organizational cybersecurity weaknesses. Our work does not suggest that organizations should suspend vulnerability assessments. Rather, it suggests the need to understand how to conduct vulnerability assessments or tests of the human element in cybersecurity in a way that does not harm employees but rather enhances their understanding of cybersecurity policies and tactics. Identifying human-centered ways to prevent betrayal while maintaining the effectiveness of vulnerability assessments could help mitigate the negative effects of vulnerability assessment demonstrated in these studies. Our post hoc moderation analysis demonstrates the importance of all forms of trust in mitigating the negative consequences of vulnerability assessments in our causal chain that can undermine the importance of human-centric cybersecurity initiatives. Furthermore, showing the contrasting amplifying and negative consequences of information security mindfulness reveals its *bright* and *dark* side effects. Further research could help guide people-centric thinking in cybersecurity research on important topics such as deception, computer misuse, artificial intelligence-assisted monitoring and surveillance, resistance (Yapo & Weiss, 2018), and associated interjections that have bright and dark side effects, as our moderation analysis shows.

8 Conclusion

Motivated by industry stories of cybersecurity failures in organizations, we designed a study to better understand employee responses to cybersecurity practices that may be perceived as detrimental to effective organizational cybersecurity. While these efforts are important and designed to protect the organization, these behaviors may lead employees to actively resist compliance with cybersecurity policies and procedures. To this end, we drew on the theory of betrayal aversion to develop a model that explains why some vulnerability assessments can lead to an affective state of cybersecurity betrayal and elicit active user resistance. We provide evidence that these tactics can induce feelings of betrayal and lead employees to resist cybersecurity. By expanding our understanding of vulnerability assessments and their implications, this study provides a foundation for future research that examines the negative impact of cybersecurity tactics on cybersecurity unit compliance with policies and technologies.

References

Accenture. (2019). *More responsible use of workforce data required to strengthen employee trust and unlock growth, according to Accenture report*. https://newsroom.accenture.com/news/more-responsible-use-of-workforce-data-required-to-strengthen-employee-trust-and-unlock-growth-according-to-accenture-report.htm

Acquisti, A., & Grossklags, J. (2005). Privacy and rationality in individual decision making. *IEEE Security & Privacy*, *3*(1), 26-33.

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, *17*(4), 351-371.

Anders, L. (2022). *How to wreck your company's culture with one bad phishing test*. Hooksecurity. https://www.hooksecurity.co/blog/how-to-wreck-your-companys-culture-with-one-bad-phishing-test

Anderson, B., Vance, A., Kirwan, C. B., Eargle, D., & Jenkins, J. L. (2016). How users perceive and respond to security messages: A NeuroIS research agenda and empirical study. *European Journal of Information Systems*, *25*(4), 364-390.

Bacudio, A. G., Yuan, X., Chu, B.-T. B., & Jones, M. (2011). An overview of penetration testing. *International Journal of Network Security & Its Applications*, *3*(6), 19.

Balozian, P., Burns, A. J., & Leidner, D. E. (2023). An adversarial dance: Toward an understanding of insiders' responses to organizational information security measures. *Journal of the Association for Information Systems*, *24*(1), 161-221.

Barlow, J. B., Warkentin, M., Ormond, D., & Dennis, A. (2018). Don't even think about it! The effects of antineutralization, informational, and normative communication on information security compliance. *Journal of the Association for Information Systems*, *19*(8), 3.

Barlow, J. B., Warkentin, M., Ormond, D., & Dennis, A. R. (2013). Don't make excuses! Discouraging neutralization to reduce IT policy violation. *Computers & Security*, *39*, 145-159.

Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, *47*(4), 644-675.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323-370.

Blythe, J., & Collins, E. (2022). *Punishment in cyber security research report*. Cybsafe. https://cybsafe-resources.s3-eu-west-1.amazonaws.com/CYBSAFE-policy+briefing-v3-20608+JH.pdf

Bordia, P., Restubog, S. L. D., Bordia, S., & Tang, R. L. (2010). Breach begets breach: Trickle-down effects of psychological contract breach on customer service. *Journal of Management*, *36*(6), 1578-1607.

Bordia, P., Restubog, S. L. D., & Tang, R. L. (2008). When employees strike back: Investigating mediating mechanisms between psychological contract breach and workplace deviance. *Journal of Applied Psychology*, *93*(5), 1104-1117.

Boss, S., Galletta, D., Lowry, P. B., Moody, G. D., & Polak, P. (2015). What do systems users have to fear? Using fear appeals to engender threats and fear that motivate protective security behaviors. *MIS Quarterly*, *39*(4), 837-864.

Boss, S. R., Kirsch, L. J., Angermeier, I., Shingler, R. A., & Boss, R. W. (2009). If someone is watching, I'll do what I'm asked: Mandatoriness, control, and information security. *European Journal of Information Systems*, *18*, 151-164.

Bulgurcu, Cavusoglu, & Benbasat. (2010). Information security policy compliance: An empirical study of rationality-based beliefs and information security awareness. *MIS Quarterly*, *34*(3), 523.

Burleson, J., Grover, V., Thatcher, J. B., & Sun, H. (2021). A representation theory perspective on the repurposing of personal technologies for work-related tasks. *Journal of the Association for Information Systems*, *22*(6), 1556-1589.

Burns, A. J., Roberts, T. L., Posey, C., & Lowry, P. B. (2019). The adaptive roles of positive and negative emotions in organizational insiders' security-based precaution taking. *Information Systems Research*, *30*(4), 1228-1247.

Burns, A. J., Roberts, T. L., Posey, C., Lowry, P. B., & Fuller, B. (2023). Going beyond deterrence: a middle-range theory of motives and controls for insider computer abuse. *Information Systems Research*, *34*(1), 342-362.

Caputo, D. D., Pfleeger, S. L., Freeman, J. D., & Johnson, M. E. (2013). Going spear phishing: Exploring embedded training and awareness. *IEEE Security & Privacy*, *12*(1), 28-38.

Cater, L., & Heikkile, M. (2021). *Your boss is watching: How AI-powered surveillance rules the workplace*. Politico https://www.politico.eu/article/ai-workplace-surveillance-facial-recognition-software-gdpr-privacy/

Cenfetelli, R. T., & Bassellier, G. (2009). Interpretation of formative measurement in information systems research. *MIS Quarterly*, *33*(4), 689-707.

Chen, A., & Karahanna, E. (2019). Life interrupted: The effects of technology-mediated work interruptions on wrok and nonwork outcomes. *Management Information Systems Quarterly*, *42*(4), 1023-1042.

Chen, Y., Galletta, D. F., Lowry, P. B., Luo, X., Moody, G. D., & Willison, R. (2021). Understanding inconsistent employee compliance with information security policies through the lens of the extended parallel process model. *Information Systems Research*, *32*(3), 1043-1065.

Chen, Y., Ramamurthy, K., & Wen, K.-W. (2012). Organizations' information security policy compliance: Stick or carrot approach? *Journal of Management Information Systems*, *29*(3), 157-188.

Chin, W. W., Thatcher, J. B., Wright, R. T., & Steel, D. (2013). Controlling for common method variance in PLS analysis: the measured latent marker variable approach. In *New perspectives in partial least squares and related methods* (pp. 231-239). Springer.

Colnago, J., Devlin, S., Oates, M., Swoopes, C., Bauer, L., Cranor, L., & Christin, N. (2018). "It's not actually that horrible": Exploring adoption of two-factor authentication at a university. *Proceedings of the Conference on Human Factors in Computing Systems*.

Conway, N., Guest, D., & Trenberth, L. (2011). Testing the differential effects of changes in psychological contract breach and fulfillment. *Journal of Vocational Behavior*, *79*(1), 267-276.

Conway, N., Kiefer, T., Hartley, J., & Briner, R. B. (2014). Doing more with less? Employee reactions to psychological contract breach via target similarity or spillover during public sector organizational change: Change and psychological contract breach. *British Journal of Management*, *25*(4), 737-754.

Costa, S. P., & Neves, P. (2017). Forgiving is good for health and performance: How forgiveness helps individuals cope with the psychological contract breach. *Journal of Vocational Behavior*, *100*, 124-136.

Coyle-Shapiro, J. A. M., & Conway, N. (2005). Exchange relationships: Examining psychological contracts and perceived organizational support. *Journal of Applied Psychology*, *90*(4), 774-781.

Coyle-Shapiro, J. A. M., Pereira Costa, S., Doden, W., & Chang, C. (2019). Psychological contracts: Past, present, and future. *Annual Review of Organizational Psychology and Organizational Behavior*, *6*, 145-169.

Craig, K., Thatcher, J. B., & Grover, V. (2019). The IT identity threat: A conceptual definition and operational measure. *Journal of Management Information Systems*, *36*(1), 259-288.

Cram, W. A., D'Arcy, J., & Proudfoot, J. G. (2019). Seeing the forest and the trees: A meta-analysis of the antecedents to information security policy compliance. *MIS Quarterly*, *43*(2), 525-554.

Cram, W. A., Proudfoot, J. G., & D'Arcy, J. (2021). When enough is enough: Investigating the antecedents and consequences of information security fatigue. *Information Systems Journal*, *31*(4), 521-549.

Crossler, R., & Posey, C. (2017). Robbing Peter to pay Paul: Surrendering privacy for security's sake in an identity ecosystem. *Journal of the Association for Information Systems*, *18*(7), 487-515.

Crossler, R. E., Johnston, A. C., Lowry, P. B., Hu, Q., Warkentin, M., & Baskerville, R. (2013). Future directions for behavioral information security research. *Computers & Security*, *32*, 90-101.

D'Arcy, J., & Herath, T. (2011). A review and analysis of deterrence theory in the IS security literature: Making sense of the disparate findings. *European Journal of Information Systems*, *20*(6), 643-658.

D'Arcy, J., Herath, T., & Shoss, M. K. (2014). Understanding employee responses to stressful information security requirements: A coping perspective. *Journal of Management Information Systems*, *31*(2), 285-318.

D'Arcy, J., Hovav, A., & Galletta, D. (2009). User awareness of security countermeasures and its impact on information systems misuse: A deterrence approach. *Information Systems Research*, *20*(1), 79-98.

Deery, S. J., Iverson, R. D., & Walsh, J. T. (2006). Toward a better understanding of psychological contract breach: A study of customer service employees. *Journal of Applied Psychology*, *91*(1), 166-175.

Deng, H., Coyle-Shapiro, J., & Yang, Q. (2018). Beyond reciprocity: A conservation of resources view on the effects of psychological contract violation on third parties. *Journal of Applied Psychology*, *103*(5), 561-577.

DeSimone, J. A., Harms, P. D., & DeSimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior*, *36*(2), 171-181.

Dhillon, G., Abdul Talib, Y. Y., & Picoto, W. N. (2020). The mediating role of psychological empowerment in information security compliance intentions. *Journal of the Association for Information Systems*, *21*(1), 152-174.

Dincelli, E., & Chengalur-Smith, I. (2020). Choose your own training adventure: Designing a gamified SETA artefact for improving information security and privacy through interactive storytelling. *European Journal of Information Systems*, *29*(6), 669-687.

Dodge Jr, R. C., Carver, C., & Ferguson, A. J. (2007). Phishing for user security awareness. *Computers & Security*, *26*(1), 73-80.

Elangovan, A. R., & Shapiro, D. L. (1998). Betrayal of trust in organizations. *The Academy of Management Review*, *23*(3), 547-566.

Feng, G., Zhu, J., Wang, N., & Liang, H. (2019). How paternalistic leadership influences IT security policy compliance: The mediating role of the social bond. *Journal of the Association for Information Systems*, *20*(11), 1650-1691.

Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *American Marketing Association*, *18*(3), 382-388.

Gefen, D., & Larsen, K. (2017). Controlling for lexical closeness in survey research: A demonstration on the technology acceptance model. *Journal of the Association for Information Systems*, *18*(10), 727-757.

George, J. F. (1996). Computer-based monitoring: Common perceptions and empirical results. *MIS Quarterly*, *20*(4), 459-480.

Gobin, R. L., & Freyd, J. J. (2009). Betrayal and revictimization: Preliminary findings. *Psychological Trauma: Theory, Research, Practice, and Policy*, *1*(3), 242-257.

Goel, S., Williams, K., & Dincelli, E. (2017). Got phished? Internet security and human vulnerability. *Journal of the Association for Information Systems*, *18*(1), 22-44.

Goo, J., Yim, M.-S., & Kim, D. J. (2014). A path to successful management of employee security compliance: An empirical study of information security climate. *IEEE Transactions on Professional Communication*, *57*(4), 286-308.

Gracia, F. J., Silla, I., Peiró, J. M., & Fortes-Ferreira, L. (2006). The state of the psychological contract and its relation with employees' psychological health. *Psicothema*, *18*(2), 256-262.

Grégoire, Y., & Fisher, R. J. (2008). Customer betrayal and retaliation: When your best customers become your worst enemies. *Journal of the Academy of Marketing Science*, *36*(2), 247-261.

Grégoire, Y., Tripp, T. M., & Legoux, R. (2009). When customer love turns into lasting hate: The effects of relationship strength and time on customer revenge and avoidance. *Journal of Marketing*, *73*(6), 18-32.

Guo, K. H., Yuan, Y., Archer, N. P., & Connelly, C. E. (2011). Understanding nonmalicious security violations in the workplace: A composite behavior model. *Journal of Management Information Systems*, *28*(2), 203-236.

Han, J., Kim, Y. J., & Kim, H. (2017). An integrative model of information security policy compliance with psychological contract: Examining a bilateral perspective. *Computers & Security*, *66*, 52-65.

Harwell, D. (2020). Managers turn to surveillance software, always-on webcams to ensure employees are (really) working from home. *Washington Post* https://www.washingtonpost.com/technology/2020/04/30/work-from-home-surveillance/

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115-135.

Herath, T., & Rao, H. R. (2009). Protection motivation and deterrence: A framework for security policy compliance in organisations. *European Journal of Information Systems*, *18*(2), 106-125.

Herath, T., Yim, M.-S., D'Arcy, J., Nam, K., & Rao, H. R. (2018). Examining employee security violations: Moral disengagement and its environmental influences. *Information Technology & People*, *31*(6), 1135-1162.

Hong, W., Chan, F. K. Y., Thong, J. Y. L., Chasalow, L. C., & Dhillon, G. (2013). A framework and guidelines for context-specific theorizing in information systems research. *Information Systems Research*, *25*(1), 111-136.

Hovav, A., & Putri, F. F. (2016). This is my device! Why should I follow your rules? Employees' compliance with BYOD security policy. *Pervasive and Mobile Computing*, *32*, 35-49.

Hsu, J. S.-C., Shih, S.-P., Hung, Y. W., & Lowry, P. B. (2015). The role of extra-role behaviors and social controls in information security policy effectiveness. *Information Systems Research*, *26*(2), 282-300.

Hu, Q., Hart, P., & Cooke, D. (2007). The role of external and internal influences on information systems security: A neo-institutional perspective. *The Journal of Strategic Information Systems*, *16*(2), 153-172.

Hwang, I., Kim, D., Kim, T., & Kim, S. (2017). Why not comply with information security? An empirical approach for the causes of non-compliance. *Online Information Review*, *41*(1), 2-18.

Jaeger, L., & Eckhardt, A. (2021). Eyes wide open: The role of situational information security awareness for security-related behaviour. *Information Systems Journal*, *31*(3), 429-472.

Jalali, M. S., Siegel, M., & Madnick, S. (2019). Decision-making and biases in cybersecurity capability development: Evidence from a simulation game experiment. *The Journal of Strategic Information Systems*, *28*(1), 66-82.

Jasso, G. (2006). Factorial survey methods for studying beliefs and judgments. *Sociological Methods & Research*, *34*(3), 334-423.

Jenkins, J. L., Anderson, B. B., Vance, A., Kirwan, C. B., & Eargle, D. (2016). More harm than good? How messages that interrupt can make us vulnerable. *Information Systems Research*, *27*(4), 880-896.

Jensen, M. L., Dinger, M., Wright, R. T., & Thatcher, J. B. (2017). Training to mitigate phishing attacks using mindfulness techniques. *Journal of Management Information Systems*, *34*(2), 597-626.

Jensen, M. L., Durcikova, A., & Wright, R. T. (2021). Using susceptibility claims to motivate behaviour change in IT security. *European Journal of Information Systems*, *30*(1), 27-45.

Jensen, M. L., Wright, R. T., Durcikova, A., & Karumbaiah, S. (2022). Improving phishing reporting using security gamification. *Journal of Management Information Systems*, *39*(3), 793-823.

Jiang, L., Probst, T. M., & Benson, W. L. (2017). Organizational context and employee reactions to psychological contract breach: A multilevel test of competing theories. *Economic and Industrial Democracy*, *38*(3), 513-534.

Johnson, J. L., & O'Leary-Kelly, A. M. (2003). The effects of psychological contract breach and organizational cynicism: Not all social exchange violations are created equal. *Journal of Organizational Behavior*, *24*(5), 627-647.

Johnston, A., Di Gangi, P., Howard, J., & Worrell, J. L. (2019). It takes a village: Understanding the collective security efficacy of employee groups. *Journal of the Association for Information Systems*, *20*(3), 186-212.

Johnston, A. C., Warkentin, M., McBride, M., & Carter, L. (2016). Dispositional and situational factors: Influences on information security policy violations. *European Journal of Information Systems*, *25*(3), 231-251.

Johnston, A. C., Warkentin, M., & Siponen, M. (2015). An enhanced fear appeal rhetorical framework: Leveraging threats to the human asset through sanctioning rhetoric. *MIS Quarterly*, *39*(1), 113-134.

Kam, H. J., Ormond, D. K., Menard, P., & Crossler, R. E. (2022). That's interesting: An examination of interest theory and self-determination in organisational cybersecurity training. *Information Systems Journal*, *32*(4), 888-926.

Karagonlar, G., Eisenberger, R., & Aselage, J. (2016). Reciprocation wary employees discount psychological contract fulfillment: Psychological contract fulfillment. *Journal of Organizational Behavior*, *37*(1), 23-40.

Kelley, C. M., Hong, K. W., Mayhorn, C. B., & Murphy-Hill, E. (2012). Something smells phishy: Exploring definitions, consequences, and reactions to phishing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.

Kelly, J. R., & Barsade, S. G. (2001). Mood and emotions in small groups and work teams. *Organizational Behavior and Human Decision Processes*, *86*(1), 99-130.

Khern-am-nuai, W., Hashim, M. J., Pinsonneault, A., Yang, W., & Li, N. (2023). Augmenting password strength meter design using the elaboration likelihood model: Evidence from randomized experiments. *Information Systems Research*, *34*(1), 157-177.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford Publications.

Kock, N., & Lynn, G. (2012). Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *Journal of the Association for Information Systems*, *13*(7).

Koehler, J. J., & Gershoff, A. D. (2003). Betrayal aversion: When agents of protection become agents of harm. *Organizational Behavior and Human Decision Processes*, *90*(2), 244-261.

Kohnke, A., Sigler, K., & Shoemaker, D. (2017). *Implementing cybersecurity: A guide to the National Institute of Standards and Technology Risk Management Framework*. CRC Press.

Krebs, S. A. (2019). *Should failing phish test be a fireable offense?* Krebs on Security. https://krebsonsecurity.com/2019/05/should-failing-phish-tests-be-a-fireable-offense/

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2010). Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology*, *10*(2), 7-31.

Langer, E. J. (1989). Minding matters: The consequences of mindlessness-mindfulness. In *Advances in experimental social psychology* (Vol. 22, pp. 137-173). Academic Press.

Liang, H., & Xue, Y. L. (2010). Understanding security behaviors in personal computer usage: A threat avoidance perspective. *Journal of the Association for Information Systems*, *11*(7), 394-413.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, *32*(1), 53-76.

Longhi, L. (2020). *GoDaddy employees were told they were getting a holiday bonus. It was actually a phishing test.* The Copper Courier. https://coppercourier.com/story/godaddy-employees-holiday-bonus-secruity-test/

Lowry, P. B., & Moody, G. D. (2015). Proposing the control-reactance compliance model (CRCM) to explain opposing motivations to comply with organisational information security policies. *Information Systems Journal*, *25*(5), 433-463.

Lowry, P. B., Posey, C., Bennett, R. J., & Roberts, T. L. (2015). Leveraging fairness and reactance theories to deter reactive computer abuse following enhanced organisational information security policies: An empirical study of the influence of counterfactual reasoning and organisational trust. *Information Systems Journal*, *25*(3), 193-273.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, *35*(2), 293-334.

Mady, A., Gupta, S., & Warkentin, M. The effects of knowledge mechanisms on employees' information security threat construal. *Information Systems Journal*, 33(4), 790-841.

Malik, K. (2023). *Are humans the weakest link in cyber security?* Astra. https://www.getastra.com/blog/security-audit/humans-in-cyber-security/#:~:text=There's%20no%20denying%20that%20humans,staff%20into%20giving%20them%20access

Marett, K. (2015). Checking the manipulation checks in information security research. *Information and Computer Security*, *23*(1), 20-30.

Mattson, T., Aurigemma, S., & Ren, J. (2023). Positively fearful: Activating the individual's HERO within to explain volitional security technology adoption. *Journal of the Association for Information Systems*, *24*(3), 664-699.

McKee, M. (2021). *People-centric security: An overdue shift in our defense paradigm.* Proofpoint. https://www.proofpoint.com/us/blog/insider-threat-management/people-centric-security-overdue-shift-our-defense-paradigm

McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, *2*(2), 1-25.

Menard, P., Bott, G. J., & Crossler, R. E. (2017). User motivations in protecting information security: Protection motivation theory versus self-determination theory. *Journal of Management Information Systems*, *34*(4), 1203-1230.

Microsoft. (2020). *Microsoft productivity score and personalized experiences: Here's what's new to Microsoft 365 in October*. https://www.microsoft.com/en-us/microsoft-365/blog/2020/10/29/productivity-score-and-personalized-experiences-heres-whats-new-to-microsoft-365-in-october/

Microsoft. (2023). *Microsoft 365 productivity illustrations.* https://learn.microsoft.com/en-us/microsoft-365/solutions/productivity-illustrations?view=o365-worldwide

Milne, G. R., & Culnan, M. J. (2004). Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *Journal of Interactive Marketing*, *18*(3), 15-29.

Mims, C. (2022). More bosses are spying on quiet quitters. It could backfire. *Wall Street Journal*. https://www.wsj.com/articles/more-bosses-are-spying-on-quiet-quitters-it-could-backfire-11663387216?utm_source=pocket_mylist

Moody, G. D., Siponen, M., & Pahnila, S. (2018). Toward a unified model of information security policy compliance. *MIS Quarterly*, *42*(1), 285-311.

Moquin, R., & Wakefield, R. L. (2016). The roles of awareness, sanctions, and ethics in software compliance. *Journal of Computer Information Systems*, *56*(3), 261-270.

Morris, J. H., & Moberg, D. J. (1994). Work organizations as contexts for trust and betrayal. In

T. R. Sarbin, R. M. Carney, & C. Eoyang (Eds.), *Citizen espionage: Studies in trust and betrayal* (pp. 163-187). Praeger.

Nehme, A., & George, J. F. (2022). Approaching IT security & avoiding threats in the smart home context. *Journal of Management Information Systems*, *39*(4), 1184-1214.

Ng, K. C., Zhang, X., Thong, J. Y. L., & Tam, K. Y. (2021). Protecting against threats to information security: An attitudinal ambivalence perspective. *Journal of Management Information Systems*, *38*(3), 732-764.

Ng, T. W. H., Feldman, D. C., & Butts, M. M. (2014). Psychological contract breaches and employee voice behaviour: The moderating effects of changes in social relationships. *European Journal of Work and Organizational Psychology*, *23*(4), 537-553.

Nguyen, C., Jensen, M., & Day, E. (2023). Learning not to take the bait: A longitudinal examination of digital training methods and overlearning on phishing susceptibility. *European Journal of Information Systems*, *32*(2), 238-262.

Nguyen, C., Jensen, M. L., Durcikova, A., & Wright, R. T. (2021). A comparison of features in a crowdsourced phishing warning system. *Information Systems Journal*, *31*(3), 473-513.

NIST. (2018). *Framework for improving critical infrastructure cybersecurity*. https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf

NIST. (2021). *NIST SP 800-115*. https://www.nist.gov/privacy-framework/nist-sp-800-115

NIST. (2022). *SP 800-53: Security and privacy controls for information systems and organizations*. https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/final

Nunnally, J. C. (1994). The assessment of reliability. *Psychometric Theory*.

Orvis, K. A., Dudley, N. M., & Cortina, J. M. (2008). Conscientiousness and reactions to psychological contract breach: A longitudinal field study. *Journal of Applied Psychology*, *93*(5), 1183-1193.

Park, E. H., Kim, J., & Park, Y. S. (2017). The role of information security learning and individual factors in disclosing patients' health information. *Computers & Security*, *65*, 64-76.

Pavlou, P. A., Liang, H., & Xue, Y. (2007). Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective. *MIS Quarterly*, *31*(1), 105-136.

PCI. (2017). PCI data security standard (PCI DDS). PCI Security Standards Council. https://www.pcisecuritystandards.org/

Pew Research Center. (2017). *What Americans know about cybersecurity*. Pew https://www.pewinternet.org/2017/03/22/what-the-public-knows-about-cybersecurity/

Pew Research Center. (2019). *Americans and digital knowledge*. https://www.pewresearch.org/internet/2019/10/09/americans-and-digital-knowledge/

Pienta, D., Tams, S., & Thatcher, J. (2020). Can trust be trusted in cybersecurity? *Proceedings of the 53rd Hawaii International Conference on System Sciences* (pp. 4264-4273).

Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *885*(879), 10-1037.

Polites, G. L., Roberts, N., & Thatcher, J. (2012). Conceptualizing models using multidimensional constructs: A review and guidelines for their use. *European Journal of Information Systems*, *21*(1), 22-48.

Posey, C., Roberts, T., Lowry, P. B., Courtney, J., & Bennett, B. (2011). Motivating the insider to protect organizational information assets: Evidence from protection motivation theory and rival explanations. *Proceedings of the Dewald Roode workshop in information systems security* (pp. 22-23).

Posey, C., Roberts, T. L., & Lowry, P. B. (2015). The impact of organizational commitment on insiders' motivation to protect organizational information assets. *Journal of Management Information Systems*, *32*(4), 179-214.

Puhakainen, P., & Siponen, M. (2010). Improving employees' compliance through information systems security training: an action research study. *MIS Quarterly*, 34(4), 757-778.

Rachman, S. (2010). Betrayal: A psychological analysis. *Behaviour Research and Therapy*, *48*(4), 304-311.

Rai, A., Constantinides, P., & Sarker, S. (2019). Editor's comments: Next-generation digital platforms: Toward human-AI hybrids. *MIS Quarterly*, *43*(1), iii-ix.

Reina, D. S., & Reina, M. L. (2006). *Trust & betrayal in the workplace: Building effective relationships in your organization*. Berrett-Koehler Publishers.

Restubog, S. L. D., Bordia, P., & Tang, R. L. (2006). Effects of psychological contract breach on

performance of IT employees: The mediating role of affective commitment. *Journal of Occupational and Organizational Psychology*, *79*(2), 299-306.

Restubog, S. L. D., Bordia, P., Tang, R. L., & Krebs, S. A. (2010). Investigating the moderating effects of leader-member exchange in the psychological contract breach-employee performance relationship: A test of two competing perspectives. *British journal of management*, *21*(2), 422-437.

Rhee, H. S., Kim, C., & Ryu, Y. U. (2009). Self-efficacy in information security: Its influence on end users' information security practice behavior. *Computers & Security*, *28*(8), 816-826.

Rivard, S., & Lapointe, L. (2012). Information technology implementers' responses to user resistance: Nature and effects. *MIS Quarterly*, 36(3) 897-920.

Robinson, S. L., & Wolfe Morrison, E. (2000). The development of psychological contract breach and violation: A longitudinal study. *Journal of Organizational Behavior*, *21*(5), 525-546.

Rosen, C. C., Chang, C.-H., Johnson, R. E., & Levy, P. E. (2009). Perceptions of the organizational context and psychological contract breach: Assessing competing perspectives. *Organizational Behavior and Human Decision Processes*, *108*(2), 202-217.

Ross, R. S. (2012). Guide for conducting risk assessments. NIST. https://www.nist.gov/publications/guide-conducting-risk-assessments

Safa, N. S., Von Solms, R., & Furnell, S. (2016). Information security policy compliance model in organizations. *Computers & Security*, *56*, 70-82.

Sarkar, S., Vance, A., Ramesh, B., Demestihas, M., & Wu, D. T. (2020). The influence of professional subculture on information security policy violations: A field study in a healthcare context. *Information Systems Research*, *31*(4), 1240-1259.

Schuetz, S. W., Benjamin Lowry, P., Pienta, D. A., & Bennett Thatcher, J. (2020a). The effectiveness of abstract versus concrete fear appeals in information security. *Journal of Management Information Systems*, *37*(3), 723-757.

Schuetz, S., Lowry, P. B., Pienta, D., & Thatcher, J. (2020b). Improving the design of information security messages by leveraging the effects of temporal distance and argument nature. *Journal of the Association for Information Systems (JAIS)*, *22*(5), 1376-1428.

Silic, M., & Lowry, P. B. (2020). Using design-science based gamification to improve organizational security training and compliance. *Journal of Management Information Systems*, *37*(1), 129-161.

Siponen, M., & Vance, A. (2010). Neutralization: New insights into the problem of employee information systems security policy violations. *MIS Quarterly*, 34(3) 487-502.

Smith, C. P., & Freyd, J. J. (2014). Institutional betrayal. *American Psychologist*, *69*(6), 575.

Söllner, M., Benbasat, I., Gefen, D., Leimeister, J. M., & Pavlou, P. A. (2016). Trust: An MIS Quarterly research curation. *MIS Quarterly*. https://www.misqresearchcurations.org/blog/2017/5/10/trust-1

Steinfeld, N. (2016). "I agree to the terms and conditions":(How) do users read privacy policies online? An eye-tracking experiment. *Computers in Human Behavior*, *55*, 992-1000.

Straub, D., Boudreau, M.-C., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the Association for Information Systems*, *13*, Article 24.

Straub, D. W., & Welke, R. J. (1998). Coping with systems risk: Security planning models for management decision making. *MIS Quarterly*, *22*(4), 441-469.

Suresh, A., Latha, S. S., Nair, P., & Radhika, N. (2014). Prediction of fight or flight response using artificial neural networks. *American Journal of Applied Sciences*, *11*(6), 912-920.

Tan, T. M., Balaji, M. S., Oikarinen, E.-L., Alatalo, S., & Salo, J. (2021). Recover from a service failure: The differential effects of brand betrayal and brand disappointment on an exclusive brand offering. *Journal of Business Research*, *123*, 126-139.

Thatcher, J. B., Wright, R. T., Sun, H., Zagenczyk, T. J., & Klein, R. (2018). Mindfulness in information technology use: definitions, distinctions, and a new measure. *MIS Quarterly*, *42*(3), 831-847.

Tittle, C. R. (2017). Refining control balance theory. In S. Henry (Ed.), *Recent developments in criminological theory* (pp. 211-244). Routledge.

Trinkle, B. S., Warkentin, M., Malimage, K., & Raddatz, N. (2021). High-risk deviant decisions: Does neutralization still play a role? *Journal of the Association for Information Systems*, *22*(3), 3.

Tsai, J. Y., Egelman, S., Cranor, L., & Acquisti, A. (2011). The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research*, *22*(2), 254-268.

Tsohou, A., Karyda, M., Kokolakis, S., & Kiountouzis, E. (2015). Managing the introduction of information security awareness programmes in

organisations. *European Journal of Information Systems*, *24*(1), 38-58.

Turnley, W. H., Bolino, M. C., Lester, S. W., & Bloodgood, J. M. (2004). The effects of psychological contract breach on union commitment. *Journal of Occupational and Organizational Psychology*, *77*(3), 421-428.

Vance, A., Jenkins, J. L., Anderson, B. B., Bjornn, D. K., & Kirwan, C. B. (2018). Tuning out security warnings: A longitudinal examination of habituation through fMRI, eye tracking, and field experiments. *MIS Quarterly*, *42*(2), 355-380.

Vance, A., Jenkins, J. L., Anderson, B. B., Brock Kirwan, C., & Bjornn, D. (2019). Improving security behavior through better security message comprehension: fMRI and eye-tracking insights. *Proceedings of the Information Systems and Neuroscience NeuroIS Retreat 2018* (pp. 11-17).

Vance, A., Lowry, P. B., & Eggett, D. L. (2015). Increasing accountability through the user interface design artifacts: A new approach to addressing the problem of access-policy violations. *MIS Quarterly*, *39*(2), 345-366.

Wagner, L. (2020). *Tribune Publishing out-evils itself with phishing email promising bonuses*. Vice. https://www.vice.com/en/article/y3z8g5/tribune-publishing-out-evils-itself-with-phishing-email-promising-bonuses

Wall, J., Lowry, P. B., & Barlow, J. B. (2015). Organizational violations of externally governed privacy and security rules: Explaining and predicting selective violations under conditions of strain and excess. *Journal of the Association for Information Systems*, *17*(1), 39-76.

Wang, J., Li, Y., & Rao, H. R. (2017). Coping responses in phishing detection: An investigation of antecedents and consequences. *Information Systems Research*, *28*(2), 378-396.

Warkentin, M., Walden, E., Johnston, A. C., & Straub, D. W. (2016). Neural correlates of protection motivation for secure IT behaviors: An fMRI examination. *Journal of the Association for Information Systems*, *17*(3), 194-215.

Weixun Li, W., Chung Man Leung, A., & Yue, W. T. (2023). Where is IT in information security? The interrelationship among IT investment, security awareness, and data breaches. *MIS Quarterly*, *47*(1), 317-342.

Williams, L. J., Hartman, N., & Cavazotte, F. (2010). Method variance and marker variables: A review and comprehensive CFA marker technique. *Organizational Research Methods*, *13*(3), 477-514.

Willison, R., Warkentin, M., & Johnston, A. C. (2018). Examining employee computer abuse intentions. Insights from justice, deterrence and neutralization perspectives: Examining the influence of disgruntlement on computer abuse intentions. *Information Systems Journal*, *28*(2), 266-293.

Wright, R. T., Campbell, D. E., Thatcher, J. B., & Roberts, N. (2012). Operationalizing multidimensional constructs in structural equation modeling: Recommendations for IS research. *Communications of the Association for Information Systems*, *30*, Article 23.

Wright, R. T., Jensen, M. L., Thatcher, J. B., Dinger, M., & Marett, K. (2014). Research note—Influence techniques in phishing attacks: An examination of vulnerability and resistance. *Information Systems Research*, *25*(2), 385-400.

Wright, R. T., & Thatcher, J. B. (2021, July 5). Phishing tests are necessary. But they don't need to be evil. *Harvard Business Review*.

Yapo, A., & Weiss, J. W. (2018). Ethical implications of bias in machine learning. *Proceedings of the Hawaii International Conference on System Sciences.*

Yazdanmehr, A., Li, Y., & Wang, J. (2022). Does stress reduce violation intention? Insights from eustress and distress processes on employee reaction to information security policies. *European Journal of Information Systems*, 32(6), 1-19.

Yazdanmehr, A., Li, Y., & Wang, J. (2023). Employee responses to information security related stress: Coping and violation intention. *Information Systems Journal*, 33(3), 598-639.

Yoo, C. W., Goo, J., & Rao, H. R. (2020). Is cybersecurity a team sport? A multilevel examination of workgroup information security effectiveness. *MIS Quarterly*, 44(2), 907-931.

Zagenczyk, T. J., Gibney, R., Few, W. T., & Scott, K. L. (2011). Psychological contracts and organizational identification: The mediating effect of perceived organizational support. *Journal of Labor Research*, *32*(3), 254-281.

Zahedi, F. M., Abbasi, A., & Chen, Y. (2015). Fake-website detection tools: Identifying elements that promote individuals' use and enhance their performance. *Journal of the Association for Information Systems*, *16*(6), 448-484.

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, *37*(2), 197-206.

Appendix A: Full Instrument

Table A1. Loadings and Cross-Loadings

Items	CFA	EFA			
		Factor 1	Factor 2	Factor 3	Factor 4
Harm					
The activities of the cybersecurity department can cause me harm.	0.91	**0.83**	0.32	0.20	-0.20
The cybersecurity department can cause me harm.	0.94	**0.84**	0.34	0.21	-0.05
The activities of the cybersecurity department can create problems for me.	0.90	**0.84**	0.30	0.17	-0.04
The cybersecurity staff can hurt me.	0.85	**0.75**	0.32	0.25	-0.07
The activities of the cybersecurity department can put me at risk.	0.87	**0.79**	0.34	0.15	-0.04
Betrayal					
Betrayed by the cybersecurity department.	0.91	0.44	**0.76**	0.25	0.18
The cybersecurity department violated my trust.	0.92	0.44	**0.78**	0.24	0.04
Cheated by the cybersecurity department.	0.92	0.41	**0.78**	0.30	0.01
Lied to by the cybersecurity department.	0.90	0.35	**0.77**	0.30	0.03
The cybersecurity department intended to take advantage of me.	0.87	0.38	**0.68**	0.40	0.04
The cybersecurity staff double-crossed me.	0.90	0.35	**0.72**	0.44	0.00
Active resistance					
Actively take steps to avoid interacting with the cybersecurity department. (Avoidance)	0.85	0.28	0.29	**0.76**	0.04
Make concentrated efforts to evade detection of the cybersecurity department. (Avoidance)	0.88	0.29	0.35	**0.72**	0.04
Try hard to keep away from the activities of the cybersecurity staff. (Avoidance)	0.85	0.25	0.36	**0.64**	0.09
End working on information technology scanned by my organization. (Abandonment)	0.89	0.22	0.32	**0.67**	0.11
Stop using information technology that could be scanned by the cybersecurity staff. (Abandonment)	0.90	0.27	0.24	**0.65**	0.04
Actively try to undermine the cybersecurity department. (Sabotage)	0.89	0.16	0.16	**0.80**	0.15
Purposefully persuade others to not use the information technology provided by the organization. (Sabotage)	0.92	0.18	0.18	**0.84**	0.06
Deliberately disrupt the activities of the cybersecurity staff. (Sabotage)	0.91	0.16	0.16	**0.83**	0.11
Self-efficacy in information security					
I feel confident in learning the method to protect my confidential information.	0.82	0.04	-0.05	-0.03	**0.82**
I feel confident using different programs to protect my confidential information.	0.83	0.09	-0.04	0.01	**0.83**
I feel confident in learning advanced skills to protect my confidential information.	0.87	0.07	-0.02	0.01	**0.86**
I feel confident using the user's guide when help is needed to protect my confidential information.	0.83	0.03	-0.02	0.02	**0.83**

Appendix B: Multidimensionality of Active Resistance

Table B1. Assessment of Dimensionality and Convergent Validity

Model	X^2	*df*	*CFI*	*RMSEA*	*SRMR*
First-Order Factor	212.495	20	0.96	0.14	0.05
Standardized Dimension	29.265	17	0.99	0.04	0.02

Table B2. Assessment of Discriminant Validity

	Unconstrained model X^2 (*df*)	Constrained model X^2 (*df*)	ΔX^2
Avoidance disassociation:			
Abandonment	6.85 (4)	81.88 (5)	75.04
Sabotage	8.19 (8)	99.23 (9)	91.03
Abandonment disassociation:			
Avoidance	6.85 (4)	81.88 (5)	75.03
Sabotage	9.77 (4)	108.23 (5)	98.52
Sabotage disassociation:			
Avoidance	8.19 (8)	99.23 (9)	91.03
Abandonment	9.77 (4)	108.23 (5)	98.52

Table B3. Superordinate Models for Active Resistance

Model	X^2	*df*	*CFI*	*RMSEA*	*SRMR*
Parallel	29.734	17	0.99	0.04	.04
Tau Equivalent	27.806	15	0.99	0.04	.04
Congeneric	20.885	13	0.99	0.04	.02

Appendix C: Scenario Examples and Expert Rankings

To verify the realism and relevance in the design of the scenarios, we took multiple steps to ensure ecological validity. First, a panel of experts comprised of seven CISOs from various industries reviewed the scenarios (Figure C1) and actions magnitude of harm (i.e., scanning email, phishing, scanning personal devices) for realism. Note that, depending on the treatment, the subjects were provided an explanation associated with the vulnerability assessment tactic. To rate the realism of each scenario, each CISO was asked to rate the agent and action scenario based on a sliding scale from 0 to 10, with 0 being *completely unrealistic* and 10 being *completely realistic* (Table C1). The scenario resulted in an average realism score of 8.0 for the actions being taken.

In addition to the rating scenarios, the CISOs were presented with a sort-and-rank exercise to determine if the variables and magnitudes of harm were appropriate. Each CISO was presented with 10 different actions that could be presented in the scenarios to assess the realism and magnitude of the action of harm (Table C2). The results showed that most CISOs ranked scanning personal devices as a high-magnitude action, phishing as a medium-magnitude action, and scanning email as a low-magnitude action.

Jamie works in a large financial institution as a money manager for high-wealth clients. To manage client investments, money managers are given access to confidential information like customer names, social security numbers, bank account numbers, tax returns, and money transfer routing numbers.

Cybersecurity is important to the organization. The organization indicates in its handbook that all computing resources are its property. Every employee is required to acknowledge they understand the policies in the handbook via signature.

Recently, it came to light that the cybersecurity department has been **< *scanning email, phishing, scanning personal devices*>** of Jamie to try to access confidential information. Jamie's **<email was scanned, was phished, personal devices were scanned>** by the cybersecurity department, and confidential information was obtained.

Email & Internet Scanning is the act of tracking, observing, and identifying vulnerabilities in the computing activities of the employee by cybersecurity on organization owned computing devices.

Phishing is the act of carefully crafting emails by cybersecurity that are sent to selected employees with the goal of convincing them to disclose network login credentials and/or download a malicious file.

Scanning Personal Devices is the act of tracking, observing, and identifying vulnerabilities in the computing activities of an employee by cybersecurity on his or her own personal computing devices (e.g., smartphone, tablet, smartwatch, etc.) that have been connected to the organization's network.

Control Condition (First two paragraphs same)
Recently, it came to light that the cybersecurity department noted Jamie's confidential information was obtained.

Cybersecurity often notifies by using intrusion detection systems (network traffic monitoring), virus scans, phishing detection, flagging harmful websites, and firewall monitoring.

Figure C1. Experiment Scenario

Table C1. CISO Scenario Realism Rating

Rater	Action
CISO 1	9.0
CISO 2	8.0
CISO 3	10.0
CISO 4	7.0
CISO 5	8.0
CISO 6	9.0
CISO 7	5.0
Overall realism rating	**8.0**

Table C2. CISO Action of Harm Factor Ranking

Actions	High	Medium	Low	Not Realistic
Scanning personal devices	4	2	0	1
Open source intelligence gathering	4	0	2	1
Keystroke logging	4	1	2	0
Phishing	3	2	2	0
Tailgating	2	3	1	1
SMS spoofing	2	3	1	1
Scanning email	2	3	2	0
Anomalous behavior detection	2	3	2	0
Phone-based social engineering	1	4	1	1
USB drop attack	1	2	4	0

Appendix D: Common Method Variance

Using a common latent method factor, we assessed common method variance following previously established guidelines in IS (Chen & Karahanna, 2019; Chin et al., 2013). First, we added the latent common method factor to the CFA (Chen & Karahanna, 2019; Podsakoff, 2003), allowing us to assess for common method bias. We compared the model fit of the CFA with and without the marker variable. The change in CFI was 0.003, which is lower than the 0.005 established standard (Little, 1997), and the change in chi-square was insignificant. Next, we assessed for common method variance on the structural paths by including the common method factor in the structural model. We compared the results of the model with and without the common method factor, noting that the significant and nonsignificant paths in the model did not change. Finally, we used the CFA marker variable technique (Chin et al., 2013; Williams et al., 2010) to detect common method variance. To use this technique, we constrained all item loadings equal to 1 and compared the change in fit. The chi-square difference was insignificant. These analyses indicate common method variance was not a threat to this research (Table D1). Common latent items are listed in Table D2.

Table D1. Common Method Bias

	Measurement model (*n* = 496)	**Measurement model with common method factor (*n* = 496)**	**Structural model (*n* = 496)**	**Structural model with common method factor (*n* = 496)**
CFI	0.97	0.97	0.96	0.95
X^2/df	395.40 / 146	464.79 / 180	810.93 / 260	916.28 / 329
RMSEA (90% Confidence Interval)	0.06 (90% C.I. 0.05-0.07)	0.06 (90% C.I. 0.05-0.06)	0.06 (90% C.I. 0.06-0.07)	0.06 (90% C.I. 0.05-0.06)
SRMR	0.04	0.04	0.08	0.08

Table D2. Latent Common Factor Items

The sky is blue on a sunny day.
Maroon 5 makes great music.
Cats are smarter than dogs.
I am not a robot.

Appendix E: Post Hoc Moderation Measures and Model

Table E1. Item and CFA

Items	CFA loading
Trust in people **(adapted from (McKnight et al., 2011)**	
The cybersecurity department acts in my best interests. (Benevolence)	0.91
The cybersecurity department does its best to help protect my confidential information. (Benevolence)	0.92
The cybersecurity department is interested in protecting me, not just the organization. (Benevolence)	0.90
The cybersecurity department is truthful in its dealings with me. (Integrity)	0.90
The cybersecurity department is honest. (Integrity)	0.89
The cybersecurity department is sincere in its protective commitments. (Integrity)	0.92
The cybersecurity department is genuine in protecting me. (Integrity)	0.93
The cybersecurity department is competent in protecting my confidential information. (Competence)	0.91
The cybersecurity department performs its role of protecting my confidential information well. (Competence)	0.92
Overall, the cybersecurity department is capable of protecting my confidential information. (Competence)	0.87
In general, the cybersecurity department is very knowledgeable about protecting my information. (Competence)	0.85
Trust in technology **(adapted from (McKnight et al., 2011)**	
The cybersecurity system is reliable in protecting me. (Reliability)	0.84
The cybersecurity system does not fail in protecting me. (Reliability)	0.86
The cybersecurity system is extremely dependable in protecting me. (Reliability)	0.90
The cybersecurity system does not stop protecting me. (Reliability)	0.78
The cybersecurity system protected me. (Functionality)	0.90
The cybersecurity system provides the protection required to do my tasks. (Functionality)	0.87
The cybersecurity system has the ability to protect my confidential information. (Functionality)	0.80
The cybersecurity system provides help for understanding how to protect my confidential information. (Helpfulness)	0.84
The cybersecurity system provided competent guidance in protecting me. (Helpfulness)	0.88
The cybersecurity system provides whatever help needed to protect my confidential information. (Helpfulness)	0.90
The cybersecurity system provides effective advice to protect my confidential information. (Helpfulness)	0.90
Trust in the organization **(adapted from (McKnight et al., 2011)**	0.90
I am totally comfortable working on information technology issued by the organization. (Situational Normality)	0.94
I feel very good about my confidential information being protected by the organization. (Situational Normality)	0.90
I always feel confident the right things happen when my confidential information is protected by the organization. (Situational Normality)	0.93
I feel that my confidential information is protected since the cybersecurity staff is backed by the organization. (Structural Assurance)	0.94
I believe that protective guarantees made by the organization make it safe to work on the organization's information technology. (Structural Assurance)	0.91
The organization helps me feel safe that my confidential information is protected by the cybersecurity staff. (Structural Assurance)	0.94
Cybersecurity mindfulness **(adapted from (Thatcher et al., 2018)**	
I am creative when protecting my confidential information.	0.74
I am often open to learning new ways to protect my confidential information.	0.74
I like to figure out new ways to protect my confidential information.	0.82
I get involved when protecting my confidential information.	0.85

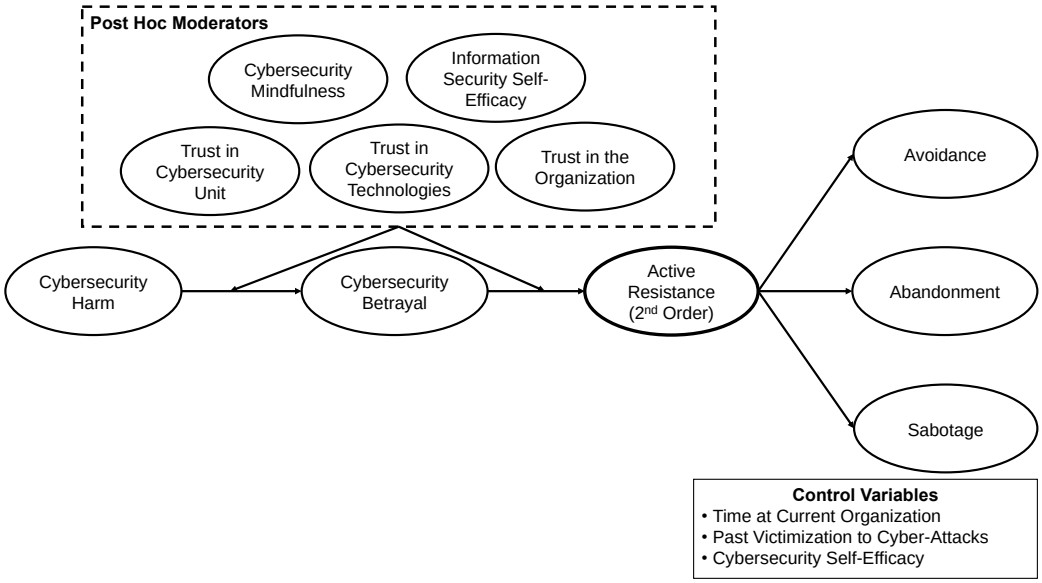


Figure E1. Mediated Moderation Analyses

About the Authors

Daniel A. Pienta is an assistant professor in the Department of Accounting and Information Management and a Research Fellow of the Neel Corporate Governance Center at the University of Tennessee, Knoxville. His research interests include information security and privacy. Before academics, he consulted some of the largest US financial institutions on due diligence and cybersecurity. Dan earned his PhD from Clemson University. He serves as an associate editor for *Journal of the Association for Information Systems*. His research has appeared in or is forthcoming in journals including the *Journal of the Association for Information Systems, MIS Quarterly*, *Information Systems Research,* and *Journal of Management Information Systems*.

Jason Bennett Thatcher is a professor in the Leeds School of Business at the University of Colorado, Boulder. He is also a full professor at the University of Manchester and an Ambassador for the Technical University of Munich. He serves as a senior editor at *Information Systems Research*, having previously served as a senior editor at *MIS Quarterly*. He has published in UT-Dallas listed journals such as *MIS Quarterly* and *Information Systems Research* and in Financial Times Top-50 journals such as *Journal of Applied Psychology* and *Journal of Management* approximately once a year since earning his PhD. He is a member of INFORMS and enjoys attending WITS. He enjoys food adventures with his zombie teen, walking his macho maltypoo, and cloudy winter days on the beach

Ryan Wright is the C. Coleman McGehee Professor and the Senior Associate Dean for Faculty and Research at the McIntire School of Commerce, University of Virginia. Professor Wright's research interests include IT security and privacy, and the diffusion of IT innovations. He has over 70 publications in outlets such as *MIS Quarterly, Information Systems Research, Journal of the Association for Information Systems*, and *Journal of Management Information Systems*. He has also garnered funding from the National Science Foundation, the State of Massachusetts, and the State of Virginia. His research has been featured in the *Harvard Business Review, The Washington Post, Forbes Magazine, USA Today,* Fast Company, and many other outlets. He has presented his research for several practitioner groups including TEDx, the Salesforce Foundation, and the Association for Finance and Technology.

Philip L. Roth is Trevillian Distinguished Professor of Management at Clemson University. His research interests involve political affiliation in organizations, employee selection, social media, and meta-analysis. Phil is a Fellow of the Society for Industrial and Organizational Psychology and the Academy of Management (AOM). He served as chair of AOM's Research Methods division and three terms as representative at large for the Human Resources division. Phil has been honored for his work in meta-analysis by the Schmidt-Hunter Meta-Analysis Award and for his contributions to the HR division with the David P. Lepak Service Award. He publishes in the *Journal of Applied Psychology, Personnel Psychology,* the *Journal of Management*, and *MIS Quarterly*. Phil earned a PhD from the University of Houston and a BA from the University of Tennessee, Knoxville.