The Data Base for Advances in Information Systems

Common Misunderstandings of Deterrence Theory in Information Systems Research and Future Research Directions

Mikko Siponen
mikko.t.siponen@jyu.fi
Wael Soliman
wael.soliman@jyu.fi
Anthony Vance
tony@anthonyvance.com

Date of Acceptance: 11/18/2020

Please cite this article as follows:
Siponen, M., Soliman, W., Vance, A. (Forthcoming). Common Misunderstandings of Deterrence Theory in Information Systems Research and Future Research Directions. *The Data Base for Advances in Information Systems*, In Press.

Common Misunderstandings of Deterrence Theory in Information Systems Research and Future Research Directions

Mikko Siponen
University of Jyvaskyla

Wael Soliman
University of Jyvaskyla

Anthony Vance
Temple University

Abstract

In the 1980s, information systems (IS) borrowed deterrence theory (DT) from the field of criminology to explain information security behaviors (or intention). Today, DT is among the most commonly used theories in IS security research. Our review of IS research applying DT highlights that many fundamental assumptions of DT are unrecognized and therefore unexamined. This may have resulted in misunderstandings and conceptual confusions regarding some of the basic concepts of DT. For example, some IS studies confuse general deterrence with specific deterrence or do not recognize the difference between the two. Moreover, these fundamental assumptions, when directly examined, may provide important information about the applicability of DT in certain IS security contexts. This research commentary aims to identify and discuss some of the fundamental assumptions of DT and their implications for IS research. By examining these assumptions, IS researchers can study the previously unexplored aspects of DT in different IS contexts. Further, by recognizing these assumptions, IS scholars can revise them and build new variants of DT to better account for specific characteristics of IS behaviors and contexts.

Keywords: Deterrence theory, deterrent effect, information security policy compliance.

Introduction

Information security issues have received increased research attention from information systems (IS) scholars, as evidenced by the special issues devoted to this subject in leading IS journals, such as *MIS Quarterly* (Mahmood et al., 2010), *European Journal of Information Systems* (Lowry et al., 2017), and *Decision Support Systems* (Luo & Zhdanov, 2016). Deterrence theory (DT) is reportedly the most frequently applied theory for explaining or predicting IS security behaviors (or intentions thereof) (D'Arcy & Herath, 2011; Lebek et al., 2014; Siponen & Vance, 2014), and it is named "the single most cited theory" (Chen et al., 2018, p. 1049) in IS security literature. During the last 30 years, DT has been applied to explain (or predict) computer abuse and misuse (D'Arcy et al., 2009; Straub, 1986, 1990; Straub & Welke, 1998), information security policy (ISP) violations by employees (Pahnila et al., 2007), Internet usage policy violations (Ugrin et al., 2008), and the illegal copying of software (Siponen et al., 2012). The ISO information security management standard (ISO17799; later renumbered as 27002) also emphasizes the use of sanctions, a key aspect of DT, to regulate employees' behaviors (Theoharidou et al., 2005).

Previous IS research has highlighted the importance of recognizing the underlying assumptions of reference theories (Grover, 2008; Truex et al., 2006). For example, Truex et al. (2006) criticized IS researchers for often borrowing theories from other disciplines "with little regard for the associated baggage of underlying assumptions" (p. 799). Similarly, Grover and Lyytinen (2015) claimed that ignoring the assumptions of theories is common in IS and note that "the review process [in IS journals] is likely to ignore the origins or justification of the proposed reference theory, its underlying assumptions, the cogency of its logic, or any controversy in the reference discipline itself" (p. 282).

In this light, DT involves many fundamental assumptions that, if recognized, will enrich our understanding of the theory and the advice IS scholars provide to practitioners. For example, an assumption of DT is that the underlying learning or explanatory mechanism is *fear-* or *pain-avoidance*. Another example of DT's fundamental assumption is that the learning mechanism for *general deterrence* is profoundly different from that for *specific deterrence* (Gibbs, 1975). Unfortunately, these and many other fundamental assumptions of DT have been overlooked by the existing DT studies in IS. This may have resulted in fundamental confusion in IS applying DT[i]. Owing to such confusions regarding the basic concepts of DT, numerous key aspects of DT have remained unexplored in IS.

In this study, we identify and critically examine eight fundamental assumptions of DT that are often unrecognized and unexplored in IS research. Unveiling these assumptions is expected to contribute to our understanding of DT and its applicability in various IS contexts. For example, failing to distinguish general deterrence from specific deterrence may hinder our understanding of which one better explains IS security behavior in a given context. Another example is the transferability effect of DT that is overlooked, owing to which we do not know how the effect of sanctions transfers from one IS security behavior to another. In yet another example, not distinguishing the mechanisms of DT as either fear- or pain-avoidance may result in some or many cases wherein the IS security behaviors are not explained by fear- or pain-avoidance.

In the cases of the above examples and others, to obtain in-depth understanding of how DT explains individuals'

security behaviors, the assumptions and mechanisms underlying DT need to be directly examined. In addition to identifying and discussing these assumptions, we propose a number of research avenues that can further our understanding of DT in the IS context.

The paper is organized as follows. In Section 2, we present an overview of DT. Next, in Section 3, we present an overview of IS security literature applying DT. In Section 4, we discuss the fundamental assumptions of DT, review their applicability in the context of information security, and present recommendations for future research directions. Finally, in Section 5, we provide our concluding remarks.

Deterrence Theory

Although many thinkers have contributed to DT, its intellectual origin can be traced back to Thomas Hobbes (1588–1679), Jeremy Bentham (1748–1832), and Cesare Beccaria (1738–1794). Hobbes believed that people pursue their self-interests. This results in conflicting interests, such as the following simplified example: I want to steal what is yours and you want to steal what is mine (Hampton, 1986). To prevent such conflicts, Hobbes argued for a social contract that can be expressed in the following manner: I give up my right to steal from you but only if you do the same, and in this way, we can maintain an organized and secure human life. According to Hobbes, the state was responsible for enforcing the social contract and the threat of sanctions should apply to all those who violate it (Hampton, 1986).

Both Bentham and Beccaria advanced the development of DT with the assumptions that people (1) have free will, (2) are pleasure-seeking, and (3) want to avoid pain (Piquero et al., 2011). Like Hobbes, Beccaria viewed people as self-interested beings who are inclined to commit crimes by definition and argued that the state needed to introduce punishment to curb the motivation to commit crimes (Paternoster, 2010). Bentham (1781) saw that people have two motives: pleasure-seeking and pain-avoidance.

Further, Beccaria and then Bentham laid down the foundation for what would later be known as the deterrent effect of punishment. They suggested that to curb the natural motivations of people to commit crime, a punishment system that outweighs their tendency for deviance must exist. Punishment, in this sense, must have three requirements: (1) sanctions must be certain, implying that there is a fear of getting caught and punished; (2) punishments must be sufficiently severe to decrease the motivation to commit crime; and (3) punishments must immediately follow the crime (Paternoster, 2010).

These early deterrence formulations have attracted much criticism from subsequent generations. In particular, between the mid-1800s and mid-1900s, several influential writers rejected DT because of its alleged distorted and simplistic representation of complex issues such as human nature and criminal motivation (Paternoster, 2010). In 1975, Gibbs reported that “most social scientists no longer considered the deterrence doctrine as deserving attention” (Gibbs, 1975, p. 9).

However, Gibbs (1975) deemed abandoning the deterrence doctrine as premature, because of a lack of empirical evidence that warranted such dismissal. In his view, the correct course of action was to transform deterrence into an empirically testable systematic theory. In transforming deterrence into such a theory, Gibbs (1975) emphasized the critical role of its underlying assumptions. According to Gibbs (1975), without such assumptions, deterrence would remain “a cogency of vague ideas” (p. 5) rather than a scientific theory. We identify a constellation of assumptions about the theory’s three core elements: (1) The punishable act (or deviancy); (2) The actor (or deviant); and (3) The punishment itself.

Regarding the *act*, because DT is mainly concerned with acts that are considered malicious or anti-social behavior (Gibbs, 1975), the prohibition of these acts must be supported by a social contract—that is, justified in the public’s eyes (Hampton, 1986). Most importantly, the prohibited act and its corresponding punishment must be clearly communicated to the public (Ball, 1955; Gibbs, 1975).

Regarding the *actor*, DT assumes that people are self-interested, egoistic beings—or as Beccaria calls them, despotic spirits—inclined to commit crime by nature (Paternoster, 2010). Further, people learn to abstain from crime either by receiving a punishment for a wrongdoing (i.e., specific deterrence) or by observing others being punished (i.e., general deterrence) (Gibbs, 1975).

Regarding the *punishment*, DT assumes that the experience of punishment is an entirely subjective matter (Gibbs, 1975), that punishment for one type of crime may deter actors from committing a different type of crime (Gibbs, 1975), and that unjustified punishment may lead to undesirable consequences, such as backfiring (Bentham, 1781).

These fundamental assumptions about DT's three core elements will constitute the basis for our discussion on the application of DT in Section 4. However, before elaborating on each of these assumptions, we present an overview of how DT has been applied in the IS literature.

Deterrence Theory in IS Literature

To understand how DT has been applied in IS literature, we conducted a comprehensive theory-oriented literature review (Rowe, 2014; Templier & Paré, 2015; Webster & Watson, 2002). Based on our stated focus on the application of DT in IS contexts, we had two clear search criteria: a) the article applies DT and b) it does so in an IS/technology context. Using the iterative process of keyword searching, backward searching, and extraction and synthesis, we obtained 84 articles in which DT has been used to varying degrees (a detailed description of the search procedure is provided in Appendix 1).

In general, our review reveals that DT has been applied in various forms in IS contexts. It has been used as the sole theoretical foundation (Higgins et al., 2005; Straub, 1990), or together with other theories (Peace, Galletta, & Thong, 2003; Workman & Gathegi, 2007), and as a source of control variables (Ifinedo, 2014; Yazdanmehr & Wang, 2016). Methodologically, most DT studies are statistical in nature (71%), mainly using the survey format (standard, factorial, or vignette-based surveys). Among the articles, only one is a field experiment that examines the impact of punishment on actual policy violation (Kim et al., 2019). Fourteen articles (17%) are conceptual or non-empirical in nature, including literature reviews (D'Arcy & Herath, 2011; Trang & Brendel, 2019; Willison, Lowry, & Paternoster, 2018), model building exercises, and research commentaries. The remaining 10 articles are almost equally divided between qualitative studies, mixed-methods studies, and simulation studies.

We also noted that the most common application of DT in IS focuses on testing the deterrent effect—that is, examining to what extent punishment or sanctions explains actors' undesirable security-related acts (or intention)[ii]. For brevity, in the following analysis, we use the term *deterrence hypothesis* to refer to arguments linking the application of punishment (or independent variables, however operationalized in the reviewed articles) with the undesirable act (or dependent variables, however operationalized in the reviewed articles)[iii]. Because the majority of empirical work is statistical in nature, we deem the deterrence hypothesis as follows: (a) supported, if the correlation between the deterrent variable(s) and outcome variable is reported as statistically significant; (b) not supported, if the correlation between the deterrent variable(s) and outcome variable is reported as statistically non-significant; or (c) partially supported, if the correlation between some, but not all, deterrent variable(s) and outcome variable is reported as statistically significant. However, in the qualitative work, the deterrence hypothesis is assessed by considering the extent to which the authors have argued that punishment has been successful in deterring the undesirable behavior.

Overall, we identified 67 empirical studies that apply and examine the deterrence hypothesis in one way or another. The deterrent effect is empirically supported in only 26 of the 67 studies (39%). In 14 studies (21%), no evidence supports the deterrence hypothesis. In the remaining 27 studies (40%), some evidence suggests partial support for the deterrence hypothesis. For example, some studies found that punishment severity, but not certainty, significantly correlates with the outcome variable (D'Arcy et al., 2009). Others found the opposite: that punishment certainty, but not severity, significantly correlates with the outcome variable (Zhang et al., 2009). Interestingly, one study showed that punishment severity significantly correlates with the outcome variable but in the opposite direction hypothesized (Herath & Rao, 2009a). That is, an increase in the perception of punishment severity is correlated with a decrease in the IS security policy compliance intentions.

Further, most studies (n=72; 86%) examined DT in the organizational context, of which, 57 articles were empirical. By considering the criminality of the studied behavior[iv] (Willison, Lowry, et al., 2018), these 57 articles may be then divided into two categories: (a) 15 studies examining the impact of sanctions on criminal (or partly criminal) behavior, and (b) 42 studies examining the impact of sanctions on non-criminal behavior (see Figure 1; a more detailed illustration is provided in Appendix 2).

Insert Figure 1 About Here

DT is applied outside the organizational context in 12 articles, 10 of which are empirical. Most of this work focuses on intellectual property issues among students—for example, the illegal copying of software (Gopal & Sanders, 1997; Siponen et al., 2012) and illegal downloading of music (Levin, Dato-on, & Manolis, 2007; Sinha & Mandel,

2008). The only exceptions are two recent studies examining the impact of governmental sanction on citizens' online behavior; in both studies, the deterrence hypothesis enjoys empirical support (Park, Kwak, & Lee, 2019; Stoycheff, Liu, Xu, & Wibowo, 2019). Interestingly, Stoycheff et al. (2019) found that governmental sanction (in the form of surveillance of online activities) not only has a deterrent effect on illegal behavior (e.g., piracy) but also has a negative impact on healthy behavior (e.g., engaging in political discussions in cyberspace).

Core Theoretical Assumptions of DT and Future Research Directions

In Section 2, we briefly introduced eight fundamental assumptions that DT holds for its three core elements (the act, the actor, and the punishment). Then, in Section 3, we offered a general overview of the IS literature applying DT. In this section, we elaborate on the eight assumptions (see Table 1), assess the extent to which these assumptions have been recognized in IS literature, and discuss the implications of this assessment for future research. In Section 4.1, we discuss the assumptions of DT and to what extend they have been examined in IS literature. Then, in Section 4.2, we offer recommendations for future research directions.

Insert Table 1 About Here

Core Assumptions

The Act

In this section, we discuss to what extent DT (1) is applicable outside of malicious acts (Gibbs, 1975), (2) assumes an underlying social contract (Hobbes in Hampton, 1986), and (3) assumes the knowledge of the act (Ball, 1955; Gibbs, 1975).

Assumption #1: Criminality of the Act

Does DT apply to all acts or only those that are malicious in nature?

One of the oldest approaches to describe the moral content of crimes in criminology is by referring to the extent to which an act is *malum in se* or *malum prohibitum* (Davis, 2006; Green, 2016). Malum in se is an act that is morally wrong "prior to and independent of its being illegal" (Green, 2016, p. 2). These acts are often seen as evil-by-nature, such as rape and murder. By contrast, in the case of malum prohibitum, "the wrongfulness of the conduct depends on the very fact of its being illegal" (Green, 2016, p. 2). In IS writings, a similar distinction is made between malicious and nonmalicious acts (e.g., information security policy violations). A malicious violation is one that is intended to cause harm (e.g., planting a malware), whereas a nonmalicious violation is one that lacks such intent (e.g., creating a simple password).

Willison et al. (2018) argued that DT was originally designed to "thwart *criminal* or *antisocial* behavior; not noncriminal noncompliance, let alone motivating good behavior" (p. 1189). Guo et al. (2011) also suggested that sanctions have no significant effect on employees' attitudes toward non-malicious violations, such as using a public Wi-Fi for business purposes or using an unauthorized storage device to carry organizational data. At first blush, this argument sounds reasonable, and out of the 84 articles reviewed, 10 articles (12%) have noted this assumption. However, only 3 of these articles used the assumption to explain the lack of (or partial) empirical support for the deterrence hypothesis (see Table 3 and Appendix 3). These studies (e.g., Willison et al., 2018) argued that DT might not be suitable to study non-criminal/non-malicious behavior[v].

Closer scrutiny raises concerns with this "non-malicious" argument. For example, this argument is contrary to empirical evidence. Namely, numerous DT studies are in the context of non-malicious violations. Yet, they found formal sanctions to be fully or partially effective (as demonstrated in Figure 1 and Appendix 2). For example, sanctions have been reported to deter cyberloafing or the non-work–related use of Internet (e.g., Ugrin et al., 2007; Pahnila et al., 2007; Peace et al., 2008). Another example is a meta-analysis (Trang & Brendel, 2019) suggesting that sanction certainty has a higher correlation with good behavior (e.g., ISP compliance) than with bad behavior (ISP non-compliance). How could these results be if sanctions (in terms of DT) are *a priori* doomed to fail as an explanatory or a predictive theory in the context of non-malicious violations? Previous DT research does not explain this.

A related question is, “Do the assumptions of DT prevent it from explaining non-criminal behavior?” Gibbs (1985) claimed that “the term deterrence applies to any act and punishment, but traditionally the deterrence doctrine is associated with criminal justice” (p. 87). Our view regarding whether DT applies to non-malicious or non-criminal behavior is the following. First, deterrence can be used for any act, including non-crimes (with certain reservations). For example, contract violations (e.g., non-disclosure agreements) in information technology (IT) or elsewhere may involve sanctions, yet the violation may not be criminal or necessarily malicious. Second, the interest of recent DT theorists such as Gibbs (1985) has been to explain criminal behavior perhaps because they are criminologists. However, the fact that criminologists examine crime and Gibbs’ (1975, 1985) DT was designed to explain crime does not mean that DT cannot be effective for explaining non-crimes. By analogy, a medical treatment originally designed for one disease is later found useful in treating another disease for which it was not originally designed (Siponen & Klaavuniemi, 2019). Similarly, many fear appeal or health psychology theories may be created for health threats. Yet, they could be effective for explaining non-health issue as well such as IS security.

Furthermore, the application of DT to ISP violations may hinge on the extent to which DT elements, such as specific and general deterrent effects, can be connected to IS security policy (ISP) violations (regardless of whether they are crimes). Consider, for example, the use of public Wi-Fi as an ISP violation (Guo et al., 2011). There may be no *deterrent effect* for using public Wi-Fi because people have not received sanctions for doing so. If this is true, the issue would be *not connecting the deterrent effect to the ISP*. In such a case, one could claim that the organization has not fully applied DT.

Moreover, organizations may have very different attitudes toward ISP violations, and the attitude may vary depending on the type of ISP violation. For example, imagine non-work–related activities of a U.S. Navy sailor (Stanko & Beckman, 2015) with a university student (Wolfe et al., 2008). Military organizations can have a very serious attitude toward non-malicious violations because they operate in a world where countries use the cyberspace to spy on one another. They have personnel whose sole responsibility is to find why non-malicious violations occur and suggest whether any punitive or similar follow-up action is required. Banks are another example of an organization where non-malicious ISP violations can be taken very seriously because of the large amounts of money involved. In contrast, a university may have a lenient attitude. In many educational institutions, the head of departments or deans could not care less if their faculty members have weak password, use USB sticks, or visit non-work–related websites at work.

Assumption #2: Social Contract Underlying the Act

To what extent does the criminalized act violate a social contract?

Closely related to the previous assumption is the assumption of a governing social contract. One might believe that an essential requirement for a functioning sanction system is the presence of a social contract (i.e., a shared understanding of the accepted rules of conduct within a social system). Hobbes, Beccaria, and Bentham have been collectively referred to as the “social contract thinkers [who] provided the foundation for modern DT in criminology” (e.g., Onwudiwe et al., 2004, p. 234), as discussed in Section 2. According to Hobbes, the state is responsible for enforcing the social contract and the threat of sanctions should apply to all those who violate that social contract (Hampton, 1986). This concept is also applicable in situations where the majority of citizens accept laws and their respective sanctions because they either ensure a safe, organized life or view them as a negotiated trade-off (Siponen, 2006).

The notion of social contract may be easily connected to malicious violations, such as the disclosure of credit card or password data by a malicious insider or hacker, because in such cases, people are likely to accept the resulting sanctions and corresponding rationales. However, for non-malicious ISP violations, such as selecting a weak password, failure to back up data, or visiting news sites during work (Internet use policies), this can be a much different case. The challenge is that employees may find it difficult to (i) justify the existence of such rules and (ii) understand how violating these rules could pose threats to their organization or society. On the contrary, employees may view non-malicious violations as harmless or even expedient (Karjalainen et al., 2013; Puhakainen & Siponen, 2010). Therefore, in cases of non-malicious violations, the social contract may not be implicitly held, especially if people are not committed to their organization’s ISPs and sanctions and instead view them as unnecessary or disruptive (Karjalainen et al., 2013; Puhakainen & Siponen, 2010). That being said, we acknowledge that the establishment of social contract may greatly vary depending on the social setting (see also assumption 1).

Interestingly, despite this being an important DT assumption, we found no trace for it being explicitly discussed in the broad range of sanction-based IS studies. Nevertheless, some efforts have been made to argue along the line that sanctions might fail to produce a desired effect if violators adhere more to conventional norms rather than the threat of sanctions (D'Arcy & Herath, 2011; Lee & Lee, 2002; Lee et al., 2004; Siponen et al., 2007; Siponen et al., 2010; Ugrin & Pearson, 2013; Workman & Gathegi, 2007). For example, Lee and Lee (2002) summoned the social bond theory—with its emphasis on attachment and commitment to referent others—to argue that deviant behavior is less explained by fear of punishment and more by "a sense of fairness" (p. 59).

For many traditional criminal acts, the social contracts are assumed as learned. This may not be the case for non-malicious ISP violations, particularly when employees do not understand how the actions listed in the ISPs protect the common good (e.g., they do not understand how a strong password is important to protect their organization). This may pose a challenge in establishing a supportive (punitive) social contract. However, even if that holds, future research should explore the implications of social contract for the deterrent effect. Is the social acceptance of sanctions for non-malicious violations linked to the extent to which sanctions are enforced, which in turn is linked to the deterrent effect? These are discussed in Section 5.

Assumption #3: Knowledge of the Act

To what extent do people know which acts are punishable?

DT assumes that subjects know what is illegal or, in the case of IS security, punishable (Gibbs, 1975). Indeed, in a review of DT, Ball (1955, p. 351) noted that "[a] law can have no deterrent influence upon a potential criminal if he is unaware of its existence." For social order, the law strives to clearly articulate what is illegal. Similarly, in the case of IS security, corporate policies or contractual agreements between two parties may define what is punishable. One might argue that citizens should be well aware of law through their upbringing. Even Bentham emphasized the clear and rationalized justification of laws. However, this assumption begs the question of whether an organization's regulations are equally well-known by its employees?

Our review shows that 18 articles (21%) have hinted to this assumption without empirically examining it (see Table 3 and Appendix 3). Collectively, these studies leverage DT to emphasize that practitioners need to establish ISPs that clearly define the appropriate and inappropriate use of IS (Harrington, 1996; Hovav & D'Arcy, 2007; Peace et al., 2003; Siponen et al., 2007; Straub, 1990; Straub & Welke, 1998; Willison, Lowry, et al., 2018). For example, Peace et al. (2003) noted that "simply having the rules on the books will do little to create change if the rules are not enforced. Punishments should be clearly defined and communicated to the relevant audience" (p. 168). Note that the knowledge assumption we revisit here goes beyond the *awareness* concept or construct, as commonly operationalized in IS research. For example, D'Arcy et al. (2009), in their survey, operationalized the awareness of ISP in terms of employees' reported level of (dis)agreement with statements of this sort: "My organization has specific guidelines that govern what employees are allowed to do with their computers." While this is an acceptable approach to measuring employees' *knowledge about* the existence of ISP (or lack thereof), it tells us little about the extent of the employees' actual *knowledge in* and understanding of the content of these ISPs, particularly of the section in an ISP that details what punishment an organization ascribes to which violation.

Considering the scarcity of IS research addressing this gap in (subjects') knowledge, we suggest that in future work, scholars make informed decisions regarding how to assess whether the subjects are aware of what is punishable or what the ISPs state.

The Actor

DT assumes that people are egoistic, self-interested beings who will learn to abstain from crime by experiencing punishment either by receiving it themselves (i.e., specific deterrence) or by observing others being punished (i.e., general deterrence) (Paternoster, 2010; Gibbs, 1975).

Assumption #4: Actors Are Self-Interested Beings

To what extent are people self-interested, and does DT only explain the behavior (or intentions) of self-interested individuals?

Some biological or pathological theories of crime view crime as the result of the criminal mind, which is only limited to certain people (Paternoster, 2010). By contrast, DT assumes that all people possess a "despotic spirit," which according to Beccaria reflects an "ubiquitous self-interest" drive to commit crime; therefore, punishment is needed to suppress this drive (Paternoster, 2010, p. 768). Similarly, Bentham assumed that individuals are motivated by two things: to seek pleasure and avoid pain (Paternoster, 2010). Hence, a key assumption of DT is that people are inherently egoistic, self-interested beings (Paternoster, 2010, p. 783). They engage in crimes; only

following self-interest will eventually lead to crime. For example, shoplifting to save money is arguably a self-interested act. However, these motivations have been doubted. For example, criminologist von Hentig (1938, p. 560) argued that such motivations are rather "unreal and simple-minded" because motives for crimes can be much more complex than just pleasure-seeking and pain-avoidance. Similarly, Toby (1964, p. 332) argued that DT is only applicable to "unsocialized" and "amoral" individuals. Accordingly, the key questions are the extent to which people are self-interested and whether sanctions deter only self-interested people, particularly in the context of IS security.

Our review reveals that this assumption has gone unacknowledged and unchallenged in DT-based research. We therefore call upon a careful re-examination of the self-interest assumption of DT in the IS security context.

Assumption #5: Actors Learn to Abstain from Deviancy via General or Specific Deterrence

Is the source of deterrent effects general or specific?

Gibbs (1975) warned that deterrence can be entirely misunderstood if we fail to distinguish between two categories of individuals (or actors): "(1) those who have suffered a punishment for having committed a crime and (2) those who have not" (p. 34). The main distinction between these two idealized actors is how they learn the deterrent effect: (a) through personal experience of punishment (specific deterrence) or (b) through observing others being punished (general deterrence). These two routes represent two radically different experiences. Gibbs (1975) argued that these two different versions of DT should be discussed. Consequently, Gibbs (1975) emphasized the importance of explicitly specifying whether one is referring to general or specific deterrence and maintained that evidence for or against one does not necessarily say anything about the other, at least as far as crimes are concerned.

General deterrence refers to the fear of punishment engendered by examples of others getting caught and punished (Gibbs, 1975). The assumption of general deterrence is that when non-offenders see the examples of offenders receiving punishments, they will learn that they could receive the same punishment if they commit the same offense. In contrast, specific deterrence refers to the idea that a personal experience of punishment will deter the offender herself or himself from committing the offence again (Gibbs, 1975). In simple terms, specific deterrence implies that an offender learns the painful consequences of offending through personal experience, whereas general deterrence implies that onlookers learn the painful consequences by observing.

Our review reveals that IS studies usually use the terms *general deterrence theory* and *deterrence theory* interchangeably without explaining why the general deterrence version was preferred over the specific deterrence version. One potential explanation for the wide use of general deterrence might be the influence of Straub's (1990) original work that uses the term general deterrence. The subsequent research perhaps used the same concept without paying much attention to the difference between general and specific deterrence. However, 16 articles (19%) seem to acknowledge this distinction. Four articles refer to the assumption to explain the non-significant results post-hoc, arguing, for example, that the deterrent effect may not be effective if there are no examples of people who have been caught and punished (Moody et al., 2018; Park et al., 2019; Skinner & Fream, 1997; Ugrin & Pearson, 2013) (see Table 3 and Appendix 3). Only one study (Kim et al., 2019) applied the assumption to the research design to empirically examine the specific deterrent effect. The results of their field experiment suggest that the employees who receive punishment for falling for a phishing attack once will learn to not fall for it in the future. Further research is needed to examine specific versus general DT.

Moreover, distinguishing between what D'Arcy and Herath (2011) called "self vs. other-referenced perceived sanction measures" (p. 650) and the assumption we discuss here (namely, actors learn to abstain from deviancy via specific or general deterrence) is important. Under "self vs. other-referenced perceived sanction measures," D'Arcy and Herath (2011, p. 650) discussed whether sanction measures refer to self or others. They introduced the concept of "more personalized sanction measures," such as "what is the chance that you would be caught/arrested/punished" (p. 650). Here, the personalized sanction measure refers to *you*. In contrast, they defined what they called "other references" as measures for sanctions when "the certainty and severity of punishment" asked for "people in general" (D'Arcy and Herath, 2011, p. 650). However, this issue of "self vs. other-referenced perceived sanction measures," albeit important, is not tantamount to the issue of how actors learn the deterrent effect.

Why is the source of the deterrent effect important, other than the fact that Gibbs' (1975) DT is based on it? Examining the deterrent effect is important to determine whether the estimated consequences are based on general or specific deterrence. If, for example, the specific deterrent effect is more effective than the general deterrent effect (or vice versa), then this information has significant ramifications for research and practice.

Examining whether the source of the deterrent effect is specific or general, as required by Gibbs (1975), would complement previous IS research.

Apart from the above example, previous research on DT in IS security suggests that employees' perceptions of sanctions are static. Provided that sanctions are learned through experience—general or specific deterrence—the learning experience with respect to sanctions does not appear to stop at a certain point in time and remains static. Rather, as advanced by Thornberry (1987), the perceptions of sanctions and their role in explaining human action change over time on the basis of experience and learning. If employees' reactions to sanctions change over time, this could explain the inconsistent results of IS security research on sanctions (D'Arcy & Herath, 2011; Lebek et al., 2014). IS scholars have commonly used one-time surveys to capture the deterrent effect. If employees' reactions to sanctions change over time, such surveys do not capture the change. This may add another layer of explanation to the inconsistent results, as different DT studies may have captured people at different stages of (learning) development. In line with the above discussion, we recommend future research to investigate how people learn and respond to both general and specific deterrent effects over time.

The punishment

Traditionally, criminologists discuss formal sanction or legal punishment, which reflects an authority's right to the "intentional inflection of suffering" (Honderich, 2005, p. 4); this suffering may be "pain, deprivation, or some other form of suffering" (Tunick, 1992, p. 1). To Gibbs (1975), "any legal action taken by legal officials is a legal punishment when it is perceived by some members of the population in question as a cause of pain or discomfort" (p. 65). For example, sanctions can induce physical pain, as in the case of flogging or whiplashing, which is still administered in some parts of the world today. Moreover, sanctions can induce emotional pain or distress, as in the case of freedom deprivation or imprisonment. The same also applies to fines, where the suffering is mostly economic. From the 1980s onward, in addition to legal sanctions, criminologists began examining extralegal or informal sanctions under DT, such as the disapproval of family or peers, moral belief, or self-imposed shame (Braithwaite, 1989; Paternoster et al., 1983), under the label of contemporary DT. However, critiques have maintained that these extralegal developments do not fit well with the basic assumptions of DT as originally envisioned, which are fear- and pain-avoidance (Akers & Sellers, 2004). In fact, Gibbs (1986, 1975) claimed that DT's sanctions refer only to legal sanctions and argued that "unless threat and fear are stressed, deterrence is a hodgepodge notion" (p. 87). A counterargument would be that fear is not only associated with legal sanctions. Williams and Hawkins (1986, p. 568–569) claimed that "the basic element of deterrence—fear—is a phenomenon that extends beyond the intrinsic elements of legal sanctions." Indeed, in health psychology, fear has been associated with life-threatening health threats that are unrelated to legal sanctions (Milne et al., 2000). Although the debate remains inconclusive, we argue that informal sanctions can be part of DT if their underlying mechanisms are fear-avoidance (Gibbs' version of DT) or pain-avoidance (Bentham's version of DT). Ultimately, this empirical question remains to be answered, and the answer may be highly contextual.

Beyond the question of how punishment deters, there exists the broader question of "why punish at all?" (Tunick, 1992). For instance, some regard that a society that cannot function without resorting to punishment, in the formal sense discussed earlier, "reflects a crisis of law and order, a crisis of legitimacy" (Tunick, 1992, p. 6). Others, by contrast, believe that the functioning of society depends to a great extent on the existence of a punitive system. However, they disagree when it comes to their justification or rationalization of the purpose of sanction. Two broad schools are often contrasted: (a) backward-looking schools/theories and (b) forward-looking schools/theories (Honderich, 2005; Tunick, 1992). Backward-looking theories place the roots of justification in the past. That is, a punishment is justified by looking back to a committed offence, regardless of the implications or consequences that temporally follow the offence. Retribution theory, or family of theories, represents one of the most popular backward-looking justifications for punishment (Honderich, 2005; Tunick, 1992). Retribution theory suggests that people may feel that justice has been served when, for example, those convicted of murder receive long prison sentences (Bean, 1982). In other words, retribution for a crime or a violation of social norms implies that violators get what they deserve. Retributivists believe not only that it is just to punish violators but also that "we have a categorical obligation to impose a certain penalty—it would be wrong not to impose it" (Honderich, 2005, p. 18).

Forward-looking theories, on the contrary, argue that "we punish for some future good" (Tunick, 1992, p. 14); therefore, they place the justification of punishment in some future societal benefit. Prevention theory, or family of theories, represents one of the most popular forward-looking justifications for punishment. In fact, Gibbs lists nine preventive aims for punishment other than deterrence, including incapacitation, enculturation, and reformation. Incapacitation suggests that the public is safer when offenders are physically removed from society (Nagin, 2013). However, as according to DT, the deterrent (preventive) effect of punishment is important. Table 2 summarizes the two main justifications for punishment.

Insert Table 2 About Here

Regarding the punishment-specific assumptions, DT assumes that experience of an objective punishment is a subjective matter (Gibbs, 1975), that punishment for one type of crime may not deter actors from committing another type of crime (Gibbs, 1975), and that unjustified punishment may lead to undesired consequences (Bentham, 1781). We elaborate on these three assumptions next.

Assumption #6: Punishment experience is a subjective interpretation

How can we formulate ISPs with the correct level of punishment if we do not know how these levels are perceived?

A key assumption of DT (Gibbs, 1975) is that the properties of punishment can only be subjectively experienced; that is, DT fundamentally relies on a subjective interpretation of an objective sanction. Gibbs models this relationship as *Lp* ←→ *Pp*, where *Lp* reflects a property of punishment (severity, certainty, or celerity) and *Pp* reflects the perception of that property by individuals. Paternoster (2010) considers this assumption to be at the core of DT: "At its core, therefore, deterrence theory is a social psychological theory of threat communication in which the causal chain runs from the objective properties of punishment through the perceptual properties of punishment to crime" (p. 785). To illustrate this distinction and relationship, Gibbs (1975, p. 8) gives an example: "If the punishment for prostitution is a $2,000 fine in one jurisdiction but a jail sentence of 30 days in another jurisdiction, which of the two punishments is more severe?" Gibbs (1975, p. 8) criticized both Bentham and Beccaria for not answering these questions.

Our review reveals that only nine articles (11%) conceptually acknowledge the objectivity–subjectivity problem without empirical examination (see Table 3 and Appendix 3). For instance, D'Arcy and Herath (2011) observed that the discrepancy in deterrence research findings may be attributed to ignoring the fact that different individuals may differently perceive the same punishment. D'Arcy and Hovav (2009) echoed this argument by noting that the perceptions of punishment vary from one individual to another and that such differences will not have a uniform effect on behavior. In a similar vein, Hua and Bapna (2013) contended that conventional punishments might not be effective in deterring cyber terrorists, who unlike typical citizens, may be willing to sacrifice their lives for a cause (p. 183). Beyond these conceptual arguments, only three articles have attempted to empirically address this assumption, where the study participants were involved in defining the perceptual level of objective sanctions (Higgins et al., 2005; Levin et al., 2007; Wolfe et al., 2008).

Following Gibbs, we view that examining this assumption (i.e., punishment experience is a subjective interpretation) is crucial in DT research in the IS context owing to three reasons. First, it is important to clarify a) which type of sanction is under question (e.g., physical, emotional, and economic; see Section 2) and b) the desired level of severity of each type of sanction in terms of the deterrent effect. Second, this assumption is particularly crucial for practical policy-making. If we do not know what level of sanctions is sufficiently severe, we cannot formulate policies with effective sanctions, as the main argument for DT is that sanctions that are not sufficiently severe do not deter; while excessively severe sanctions could backfire (see Assumption #8). Thus, examining which types of sanctions have the most effective deterrent effect provides practitioners the necessary information to determine the sanctions they want to use. Third, the method of measuring severity currently adopted in most IS research is problematic. In particular, severity is commonly operationalized in the form of asking subjects to declare the extent to which they agree or disagree with statements such as, "If I do X, I would have severe sanctions," which does not specify what constitutes a severe sanction. Consequently, we do not know to what extent severe sanctions are the same for everyone, and future research may take various routes for examining the perceived level of sanctions severity.

Assumption #7: Punishment for One Act May Deter Another

To what extent is the deterrent effect of punishment for one crime transferable to another?

Gibbs (1975, p. 37) raised the issue of transferability—that is, the extent to which a punishment experience (general or specific) for one crime is transferable to other crimes. Subsequently, Gibbs (1986) noted the conundrum that the extent to which "punishment for certain type of crime deters one not to engage in another type of crime is not yet solved." In the IS literature applying DT, the extent to which an experienced sanction or

punishment is transferable within or between different types of IS security acts is entirely unexamined.

We encourage that future IS research provides new insights regarding sanction transferability, a question first raised by Gibbs (1975) and which remains to be answered in ISP violation contexts. Considering the paucity of research in this area, IS researchers could investigate the extent to which receiving a sanction for one ISP violation deters users from committing a different violation in the future.

Assumption #8: Punishment May Produce Undesired Effect

To what extent do sanctions actually lead to the anticipated effect?

The legal system in the pre-enlightenment era was seen by many as cruel and barbaric, since it allowed horrific practices such as torture, secret accusations and even convictions without trials. These practices provoked enlightenment thinkers, such as Beccaria, to propose radical changes in the fundamental ways authority perceived and applied punishment. For instance, Beccaria proposed legal reforms that mandated clearly written laws, forbidding all forms of torture, and above instating the proportionality between crime and its corresponding punishment. In Beccaria's view, not only are these reforms just, but also rational. (Paternoster, 2010). Indeed, rationally speaking, disproportionate and unjustified punishments may produce undesired consequences. Bentham argued that sanctions having a weak or lacking social contract (see assumption 2) may backfire because they are considered unwarranted. Bentham (1781) posited that excessive sanctions are unjust and can backfire with negative consequences. Quinlan (2010) noted that the punishments perceived as unjust or unfair can in fact increase, not decrease, the likelihood of re-offending (i.e., specific deterrence in reverse, see assumption 5). Some crime intervention programs (Petrosino et al., 2000) have actually led to more crimes (Sherman et al., 1998). As Petrosino et al. (2000) put it, "despite our best intentions, [such] programs can not only fail to reach objectives but can backfire, leading to more harm than good" (p. 371).

Our review reveals that 12 articles (14%) have acknowledged this assumption, five of which refer to this assumption to explain the empirical negative consequences of sanctions (see Table 3 and Appendix 3). However, none of these studies were designed to study the backfiring effect. The main argument of these studies is that regardless of whether the deterrence hypothesis is empirically supported, readers should be cautioned to not "over-do it" (Chen & Li, 2014; Dhillon et al., 2004; Herath & Rao, 2009a; Knapp et al., 2007; Lowry et al., 2015; Sinha & Mandel, 2008). One of the first IS studies to highlight this problem was Dhillon et al.'s (2004) case study on computer-related crime in a United Nations' medical research center. The study showed that excessive controls following criminal acts disrupted the organization's activities because employees in the research center (though not involved in the crime) felt alienated; consequently, most research projects failed to finish on time and ended up being over budget.

In a similar vein, Lowry et al. (2015) warned organizations from strongly exerting information security controls on their employees, which can cause the employees to react in undesirable ways. Further empirical support for this dilemma comes from Herath and Rao's (2009a) and Rajab and Eydgahi's (2019) surprising finding on the role of sanctions on policy compliance intentions. Although Herath and Rao (2009a) hypothesized and expected that severe sanctions would positively influence compliance intentions, the results showed that sanctions severity decreased the likelihood of compliance intentions. Rajab and Eydgahi's (2019) findings suggested that as the probability of detection increased, the intentions of employees' to comply with ISP decreased.

In general, we note that this assumption of DT has been largely unexplored by IS security researchers. The negative consequences of sanctions could include decreased work productivity, motivation, trust, organizational commitment, and security attitudes, as individual employees will react to the imposition of sanctions that they feel are disproportionate (Karjalainen et al., 2013). Further research is required to examine these consequences. Table 3 lists the assumptions and the articles in which they are conceptually recognized and empirically applied.

Insert Table 3 About Here

Future Research Directions

In Section 4.1, we revisited the eight fundamental assumptions of DT related to its three core elements: the act, the actor, and the punishment. Our discussion of these assumptions highlights the issues that have largely gone unnoticed in the IS literature. In this section, we turn to some general as well as specific research directions that these insights point to.

Regarding the general research directions, of the 84 articles addressing DT, only four are qualitative in nature. Qualitative research (e.g., case study, ethnography, and phenomenology) may add a layer of understanding to the issues that cannot be easily captured via statistical research settings. While statistical research especially in the hypothetico-inductive-statistical settings allows data, observation or phenomenon inspired hypothesis (Siponen & Klaavuniemi, 2020), qualitative research is also important in understanding the IS security context specifics aspects of DT. Qualitative research is also useful to capture change and dynamics. Regarding the specific research directions, we now discuss how each of the eight assumptions could be translated into actionable research directions (RDs). Note that we do not expect future research to test all the stated assumptions in a single study, which is in line with Gibbs' (1975) DT.

RD #1: The Relevance of the Criminality of the Act in ISP Violations

A previous study suggested that DT is designed for "criminal/malicious behaviors" and that it must be "carefully recontextualized" in a different context, such as non-malicious violations (Willison et al., 2018, p. 1189).[vi] However, empirical findings in IS challenge the views that DT only apply in malicious/criminal behaviors. Moreover, the fact that DT was proposed by criminologists to apply to crimes does not *necessarily* mean it is ineffective in other contexts. Several examples exist in sciences where theories, models, and treatments are later found useful in different context. For example, Protection Motivation Theory was originally proposed for health threats, yet it has been found useful in other contexts, too (Westcott et al., 2017).

We emphasize that DT may be ineffective in ISP violations because some key DT assumptions are not considered in the context of ISP violations. Consider, for example, a case where organizations do not give sanctions for ISP violations, such as not locking computers, and employees became aware of them. This lack of sanction (rather than the non-criminal nature) can theoretically explain the ineffectiveness of sanctions in terms of DT (see assumption 5). Before a new contextualized DT can be built, knowing to what extent the many existing DT assumptions apply in different IS security contexts is important. These assumptions are presented as assumptions 1–8.

Moreover, a clear-cut separation between non-malicious and malicious acts can be misleading because the same act can be malicious or non-malicious based on a person's intent. One can share a password with non-malicious or malicious intent. Also, spreading ransomware is harmful irrespective of whether the employee clicks a phishing link with non-malicious or malicious intent. Rather than malicious or non-malicious acts, we suggest the concept of perceived severity of the violation as potentially more important in IS applications of DT. The underlying rationale is that organizations can have different attitudes to so-called non-malicious violations.

In other words, the same act can have different levels of severity in different organizations. Browsing the web in a university environment for non-work purposes may be regarded as non-serious and non-harmful. However, that same act in a banking, military, or industrial power environment can be regarded a severe violation. Our proposition is that the perceived severity of the act can result in different outcomes in the deterrent effect or the effectiveness of sanctions. Moreover, the negative consequences and unjustness of sanctions (see assumption 8) can be less when the perceived severity of the act is high. Specific hypotheses could be that that increased perceived severity of the act increases the probability of effectiveness of the deterrent effect while decreasing the probability of sanctions' negative consequences.

If these propositions are supported by future empirical studies, the role of perceived severity of the act can be fundamental in DT applications in the IS context. However, before testing these propositions, measuring the perceived severity of an ISP violation is necessary. In the absence of better measures, scholars can start with measures such as "I regard the [VIOLATION TYPE] as serious." However, more sophisticated measures for the *perceived severity of an ISP violation* are needed.

Insert Figure 2 About Here

Developing measures for the *perceived severity of an ISP violation* is more than just a simple instrument development process. This is because what employees' perceptions of severe may depend on various sources such as IS security competence and IS security culture, which may also lack proper measures (see Figure 2). Development of such measures (e.g., for IS security competence and IS security culture in an organization) would contribute to research beyond DT because, for example, the IS security competence level can be relevant in many other theories. Such research would extend DT by adding how the deterrent effect can not only vary with

respect to organizational IS security culture and perceived severity of an ISP violation but also be dynamic (see also assumption 5). In this regard, research approaches that attempt to capture the fluctuations in perceptions and behaviors are welcome (D'Arcy & Lowry, 2019).

Besides *perceived severity of an ISP violation*, the interpretation of ISP may vary from one organization to another. For example, different organizations—from military to university—could have somewhat similar ISPs, substance-wise. To be more specific, they could have a norm in their IS policies, such as "do not use USB sticks" or "do not reuse passwords" (albeit not in these precise words). However, even if the norms are similar, their interpretation can be different. Wittgenstein (1953), among others, argued that any norm requires interpretation. How seriously the norm in the ISP guideline is interpreted can be different for employees in different types of organizations. For example, in military and bank settings, the norm "do not use USB sticks" can be interpreted more to-the-letter and without (or with less) exceptions; whereas in more relaxed environments, such as universities, the interpretation of the same norm could be closer to *supererogation* (adopting the terminology by Siponen & Iivari, 2006). Supererogation means that a norm is regarded as a highly virtuous IS security act, but it is not required in the sense that its non-occurrence would not be regarded as blameworthy. Following this concept, future researchers should theorize and explore whether ISPs and guidelines are viewed as deontologically as Kant's categorical imperatives or the Ten Commandments in the Bible (i.e., to-the-letter and absolutely), or are they regarded more flexibly as supererogatory matter? The proposition is that regarding ISPs as more flexible or supererogative may partially explain why the sanctions do not lead to the intention to comply with the ISP or compliance behavior (Figure 3.1). Also, the interpretation of the ISP may influence the perceived severity of its violation (Figure 3.2).

Insert Figures 3.1 and 3.2 About Here

RD #2: The influence of the social contract underlying the (punishable) act

In contexts wherein there is a social contract regarding certain behaviors, such as theft or murder, informal sanctions (e.g., from peers or family members toward the behavior) may ensue because the community or society disapproves of the behaviors. In contrast, if there is no social contract (or if a social contract does not demonize a certain behavior) and the community influencing the person (e.g., peers, family members, media, and superiors) does not disapprove of non-malicious ISP violations (e.g., visiting non-work–related websites on work computers), informal sanctions are difficult to develop and formal sanction may backfire (see assumption 8). If establishing a social contract is essential for deterring less serious (or non-malicious) violations (that we do not know yet), examining whether employees can be persuaded to accept social contracts for the ISP violations in their organizations is necessary. Research on the establishment of social contracts could begin with educational interventions, starting with an organization's top management, to foster a culture of disapproval that could be enforced by informal sanctions for certain types of ISP violations. Such an action research study could have the objective of ascertaining how different educational strategies and argumentations can be used to discourage undesirable information security behavior.

Case studies or action research could be used to examine how social contract for ISP violations is created or can be created. Various social and political doctrines such as Rawl's (1971) theory of justice could be used to create the social contract. Stage models such as the one illustrated in Figure 4 can be used for this purpose.

Insert Figure 4 About Here

When some measures for the social contract exist, we can examine to what extent the lack of a social contract for non-malicious ISP violations is linked to the ineffectiveness of sanctions (see Appendix 2). This issue may be linked with the perceived severity of the ISP violation (see RD #1). Thus, the severity of the violation may help in creating a social contract for sanctions. In this way, the creation of a social contract can be implicit or a side product of perceived severity of the violation. In turn, this may, result in propositions such as the perceived severity of the ISP violation increases the probability of creating a social contract in favor for sanctions associated

with ISP violations (see Figures 5.1 and 5.2).

Insert Figures 5.1 and 5.2 About Here

Alternative approaches could be also developed for creating a social contract. One approach could be campaigning. In particular, social contract could be created by persuasive messages, which could be designed in a simple two-condition experimental design:

Experimental group: pretest → social contract message → sanction notification → post-test

Control group: pretest → sanction notification → post-test

RD #3: The influence of knowledge of the (punishable) act on ISP violations

A key issue in the application of DT is to ensure that employees are aware of what the ISPs state and which ISP violations are punishable. To achieve this, IS security scholars could develop tests that go beyond measuring employees' awareness of the existence of ISPs and measure their awareness of the actual content of specific policies. As an illustrative example, such measures could include multiple choice questions regarding the context of the policy and could require employees to recall something from the ISPs and write it down. Such tests can reveal the different categories of employees: (i) those who know most ISPs, (ii) those who know some ISPs, and (iii) those who do not seem to know any ISPs. (These categories are just examples.) While knowledge of ISPs could be a moderator between sanction constructs and intention/behavior, separately applying DT models for each category might offer interesting results that require further theorizing and research. For example, DT elements could work differently with different groups.

RD #4: The role of self-interest in DT and its relation to IS security

As mentioned earlier, according to DT, people are inherently self-interested, pleasure-seeking actors with little concern for others: an assumption that has raised numerous doubts with IS researchers but has not been investigated (Toby, 1964; von Hentig, 1938). We see this (whether all people really seek pleasure and avoid pain and whether DT only applies to a limited number of people) as an empirical issue requiring investigation in future research. We invite future research to explore whether sanctions deter primary self-interested people. To this end, some measures of self-interest are needed. Although different measures of self-interest can be used, a promising direction is to use Kohlberg's (1981) cognitive theory of moral development, in which the first stage reflects egoistic interests. Instruments to test Kohlberg's (1981) cognitive theory of moral development already exist.

Alternatively, instrumentation to determine the level of employees' moral development in the context of ISP violations can be developed. The most well-known theory of moral development by Kohlberg (1981) was based on Heinz' dilemma. Heinz' dilemma is a scenario of stealing. Existing IS security scenarios can be use (and modify if needed) to develop a similar ethical dilemma.

RD #5: Whether actors learn to abstain from deviancy via general or specific deterrence

IS literature applying DT has not examined the important link between the source of the deterrent effect (specific or general) and the estimated consequences (Figure 6). IS studies typically measure the *estimated perception of sanctions*, whereas the *source of the deterrent effect* remains unmeasured. The following two examples highlight this. Siponen and Vance (2010) measured employees' estimates perception of sanction certainty using items such as "What is the chance you would receive sanctions if you violated the company ISP?" Herath and Rao (2009a) examined sanction severity using items such as "My organization terminates those employees' work contracts who repeatedly violate the organization's ISP." Both approaches are appropriate. However, these approaches (or other approaches in IS) do not examine how the deterrent effect is learned (Gibbs, 1975). Thus, in IS literature, the link between the source of the deterrent effect (specific or general) and the estimated consequences is missing.

Insert Figure 6 About Here

Many existing DT studies can bridge this research gap in Figure 6 by adding *general* or *specific deterrence* experience to existing DT models in IS. We suggest that future researchers examine this issue by making a clear

distinction between the sources of the deterrent effect. As noted earlier, IS researchers have customarily focused only on the estimated perception of sanctions in the dimension of deterrence (i.e., a sanction and its properties; Figure 6). Moreover, existing research has not examined whether the deterrent effect is specific or general (with the exception of Kim et al., 2019). As previously suggested, future researchers could approach this challenge by distinguishing between specific and general deterrence in line with the following illustrative example:

Specific deterrence: "I have received a sanction (e.g., a three-day salary deduction) for doing act X." **General deterrence**: "I have seen others receive a sanction (e.g., a three-day salary deduction) for doing act X."

The following figure (Figure 7) illustrates this using only sanctions severity. Sanctions certainty and celerity can also be added to the figure but are omitted for simplification.

Insert Figure 7 About Here

Moreover, the specific and general sanctions may vary depending on the type of ISP violation. Therefore, future research should examine which types of sanctions are more relevant or effective in IS security contexts. For example, general deterrence may be effective for cases that have few violations, whereas specific deterrence may be needed for cases that more rarely occur. However, this is merely a speculation. Indeed, there may not be a guiding theory for explaining this question. Rather, there is a need to empirically explore this issue using explorative approaches. For general deterrence, more specific issues merit future research. It may make a difference as to who the *other* people receiving sanctions are. For example, are the people who are closer to you or whom you trust more important sources of general deterrence than those who are not so close or are less trustworthy? Another important but unexplored issue is whether general deterrence is based on observing or hearing about sanctions from others. For example, hearsay could be less effective to invoke the general deterrent effect than personal observation of a colleague being punished.

RD #6: The influence of subjectiveness of the punishment

One of DT's fundamental assumptions is that unless sanctions are sufficiently severe, there is no deterrence. To precisely define severe sanctions, IS researchers must develop measures to capture how different people perceive sanctions. Very few studies have considered the empirical implications for this assumption (Higgins et al., 2005; Levin et al., 2007; Wolfe et al., 2008). Levin et al.'s (2007) work on illegal music downloading provides a one example in this direction. They developed their manipulation of severity based on an accumulation of their participants' subjective interpretations of severity. This procedure led them to identify three levels of severity that are co-developed with their study participants: (1) weak (required to delete all music files from your computer), (2) moderate (will have to perform community service of up to 25 h), and (3) strong (for each song, you will have to pay a fine of $2,500).

Closely related, scenario-based studies or field experiments could use message manipulations that present a range of specific sanctions (Jasso, 2006). Moreover, qualitative or interpretive interview studies can increase our understanding of what sanction severity implies and how people perceive the *just* levels of sanctions for different ISP violations. Such studies can also explore the sanctions that are considered severe by actors in the studied context as well as their impact as deterrents. The qualitative studies could lead to the development of instrumentation that better captures the subjective nature of sanction severity. A critical direction for future research is to explore what is severe (enough) for employees for different ISP violations because organizations (following Bentham's DT) should not give sanctions that are more severe than needed.

RD #7: Punishment for one type of action may deter a different type of action

This assumption has been entirely overlooked, despite its important implications. Future research should explore whether the deterrent effect is transferable in IS security contexts. This research direction should take into consideration the distinction made earlier regarding the source of the deterrent effect. The transferability of specific deterrence means the extent to which having personally experienced a sanction S for violation X is transferable to other IS security violations (e.g., Y and Z). In contrast, the transferability of general deterrence means the extent to which having observed someone receiving a sanction S for violation X is transferable to other IS security violations (e.g., Y and Z).

Table 4 clarifies the distinction between the transferability and learning assumptions (i.e., offenders learn via general and specific deterrence; see assumption #5).

Insert Table 4 About Here

To further clarify, let us presume a ransomware incident has occurred owing to a person clicking a phishing link and that person has received a sanction (whether formal or informal). *Violations of different nature* refer to the issue of how sanction experience transfers to different IS security actions, such as strengthening employees' passwords or encrypting confidential emails.

Violations of similar nature refer to how, for example, in the phishing scenario above, the deterrent effect would transfer to improved USB-stick usage behavior, which could be seen as similar to clicking a link (both can be sources of malware). However, what is considered similar in nature by the employees is itself not known and must be empirically studied. Such studies would clarify which information security acts are perceived as similar to one another.

The answers to these questions could be linked to how important information security is regarded in terms of the perceived severity of an ISP violation (see RD #1). For example, the experience could transfer if the severity of an ISP violation is perceived as high (see Figure 8).

Insert Figure 8 About Here

What perceived severity of an ISP violation is considered sufficient in the case of the transferability effect requires further exploration. Interviews, lab and field experiments, case studies, or scenario studies could be used to examine this question.

RD #8: How sanctions can backfire

The possibility of sanctions backfiring is a recognized DT assumption, which is not yet systematically studied in IS studies applying DT. As mentioned, backfiring can be related to IS security action (sanctions do not deter but increase ISP violation intention or behavior). However, backfiring could also be related to non-IS security actions and can manifest itself as hampered loyalty, decreased organizational commitment, revenge, and so on. Thus, examining the potential effects of backfiring, such as work satisfaction and job performance, is important. Such negative consequences, according to Bentham, mainly occur when sanctions are seen as unjustified. Case studies, interviews, and scenario settings can be used to explore these consequences. As an example of specific research direction, one could investigate the potential negative consequences of unjustified sanctions, such as decrease in organizational commitment and IS security attitudes (see Figure 9).

Insert Figure 9 About Here

The baseline model can be extended or follow-up studies can be designed. Instead of just sanctions in Figure 9, the models can vary severity, certainty, and celerity. Moreover, the experience of general or specific sanctions can be studied or measured when applicable. The baseline models can be further extended, for example, to examine the impact of sanction backfire on job satisfaction or job performance (modeled as "Impact on job") or on IS security intention or attitude (see Figure 10).

Insert Figure 10 About Here

If unjustified sanctions increase the likelihood of decreased organizational commitment or IS security attitude or they have an impact on job, the level of perceived justifiability of sanctions that lead to the backfiring effect needs

to be measured. This information can then be used as a basis for running models such as Figures 9 and 10.

Moreover, the backfiring effect as described above could be associated with the learning assumption (see assumption 5); that is, when individuals experience certain events over time, their reactions are dynamic. They might move backward according to the *stages* of the deterrent effect. To develop such a dynamic view of deterrent effect development, process and stage theories would be useful (Tsohou et al., 2020; Weinstein et al., 1998).

In summary, future research applying DT could benefit from considering our proposed research directions in the broader or more specific sense. A final note: although Gibbs (1975) endeavored to outline DT as a systematic theory, he deemed a number of questions empirical, such as the transferability effect. In other words, according to DT by Gibbs (1975), there can be a transferability effect from sanctioning act X to some other acts Y or Z, but DT does not say (or "theorize") which acts have such a transferability effect. As far as Gibbs (1975) is concerned, empirics will show it. Putting the same in the IS security context, there could be a transferability effect from one ISP violation to another, but there is currently no theory as far as DT is concerned that clarifies the transferability effect. Our point is similar to that by Gibbs (1975): some of these research issues are currently empirical or explanatory, and the theoretical reasons may be developed later based on the findings.

Conclusion

DT is among the most frequently applied theories in IS security. Despite its wide influence, numerous key assumptions of DT have remained unexplored in IS. Moreover, many fundamental concepts of DT are misunderstood or confused in IS. In this study, we identified and clarified these assumptions via answering eight questions: (1) Does DT apply to all acts or only those that are malicious in nature? (2) To what extent does the criminalized act violate a social contract? (3) To what extent do people know which acts are punishable? (4) To what extent are people self-interested, and does DT only explain the behavior (or intentions) of self-interested individuals? (5) Is the source of deterrent effects general or specific? (6) How can we formulate ISPs with the correct level of punishment if we do not know how these levels are perceived? (7) To what extent is the deterrent effect of punishment for one crime transferable to another? (8) To what extent do sanctions actually lead to the anticipated effect? Not only does answering these questions clarify many of the ambiguities of deterrence theory, but also it points to unexamined research issues that merit the attention of future DT research in the IS context. Finally, we believe that our work will help information security practitioners make better informed decisions about the application of sanction, its justifiability, and potential consequences.

Notes

[i] For example, Boss et al. (2015) report "a shift from earlier General Deterrence Theory (GDT)-based approaches to a stronger emphasis on PMT" (p. 838). They maintain that a "key reason for this shift is that GDT and RCT are based on a foundation of command and control, whereas PMT is based on the idea of using persuasive messages that warn of a personal threat" (ibid p. 838). However, Gibbs (1975) suggests that the mechanism of DT is fear avoidance, not "command and control." Moreover, DT is not the same as GDT. DT has two different variants, the general deterrence theory (GDT) and specific deterrence theory, which are commonly confused in IS. Finally, the application of DT can be "persuasive messages that warn of a personal threat" (Boss et al., 2015, p. 838) of sanctions.

[ii] Note that a considerable number of studies applied DT to examine the role of punishment on enforcing good behavior, such as policy compliance and protective behavior, rather than the deterrent effect on deviancy.

[iii] Review articles on DT in the IS context, especially by D'Arcy and Herath (2011) and Trang and Brendel (2019), have focused on inconsistent findings. We therefore do not discuss such inconsistencies in this study.

[iv] It should be noted that this classification scheme of sanction as criminal/malicious-vs.- noncriminal/nonmalicious behavior is an idealization and does not necessarily reflect the true nature of security violations. This classification approach follows the convention in IS research to classify security violations based on the available information about the security violation and its operationalization in the original article (see, e.g., Trang & Brendel, 2019; Willison et al., 2018). For instance, if the original study frames the security violation as something clearly criminal, such as, 'stealing and selling organizational information' (Hu et al., 2011), then the study is classified as criminal/malicious. By contrast, if the security violation in question does not reflect criminality/maliciousness, such as 'cyber slacking' (Ugrin et al., 2008), then the study is classified as nonmalicious/noncriminal. Conversely, a study is classified as partly criminal/malicious if it combines both types of violations (e.g., Lee et al., 2004).

[v] Regarding non-malicious ISP violations versus malicious violations (Willison & Warketin, 2013), non-malicious ISP violations may be volitional or unintentional. An example of the latter is forgetting to back up data (Malimage & Warkentin, 2010), and an example of the former is sharing passwords (Siponen & Vance, 2010). The intentional theft of intellectual property or destruction of data (e.g., passwords or credit card data) by an insider are examples of malicious violations (Willison & Warkentin, 2013).

[vi] "…although the purpose of DT is to explain how to deter criminal/malicious behaviors, several studies adopted the opposite approach and used DT to predict good behaviors" (Willison et al., 2018, p. 1190).

References

Akers, R. L., & Sellers, C. S. (2004). *Criminology Theories: Introduction, Evaluation, and Application*. Los Angeles, CA: Roxbury Press.

Albery, I. P., & Guppy, A. (1995). Drivers' differential perceptions of legal and safe driving consumption. *Addiction*, *90*(2), 245–254. https://doi.org/10.1046/j.1360-0443.1995.90224510.x

Alshare, K. A., Lane, P. L., & Lane, M. R. (2018). Information security policy compliance: a higher education case study. *Information and Computer Security*, *26*(1), 91–108.

Ball, J. C. (1955). The deterrence concept in criminology and law. *Journal of Criminal Law, Criminology and Police Science*, *46*(3), 347–354. https://doi.org/10.2307/1139417

Balozian, P., & Leidner, D. (2017). Review of IS security policy compliance: Toward the building blocks of an IS security theory. *Data Base for Advances in Information Systems*, *48*(3), 11–43.

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

Bean, P. (1982). *Punishment - A Philosophical and Criminological Inquiry*. John Wiley & Sons, Incorporated.

Beaudry, A., & Pinsonneault, A. (2010). The other side of acceptance: Studying the direct and indirect effects of emotions on information technology use. *MIS Quarterly*, *34*(4), 689–710.

Bentham, J. (1781). *An introduction to the principles of morals and legislation*.

Braithwaite, J. (1989). *Crime, Shame and Reintegration*. Cambridge, UK: Cambridge University Press.

Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2010). Information security policy compliance: An empirical study of rationality-based beliefs and information security awareness. *MIS Quarterly*, *34*(3), 523-A7.

Burns, A. J., Posey, C., Courtney, J. F., Roberts, T. L., & Nanayakkara, P. (2017). Organizational information security as a complex adaptive system: Insights from three agent-based models. *Information Systems Frontiers*, *19*(3), 509–524.

Chen, H., & Li, W. (2014). Understanding organization employee's information security omission behavior: An integrated model of social norm and deterrence. *Pacific Asia Conference on Information Systems*. Retrieved from http://aisel.aisnet.org/pacis2014/280

Chen, X., Wu, D., Chen, L., & Teng, J. K. L. (2018). Sanction severity and employees' information security policy compliance: Investigating mediating, moderating, and control variables. *Information and Management*, *55*(8), 1049–1060.

Chen, Y., Ramamurthy, K. (Ram), & Wen, K.-W. (2012). Organizations' information security policy compliance: Stick or carrot approach? *Journal of Management Information Systems*, *29*(3), 157–188. https://doi.org/10.2753/MIS0742-1222290305

Cheng, L., Li, Y., Li, W., Holm, E., & Zhai, Q. (2013). Understanding the violation of IS security policy in organizations: An integrated model based on social control and deterrence theory. *Computers and Security*, *39*, 447–459. https://doi.org/10.1016/j.cose.2013.09.009

Choi, M., & Song, J. (2018). Social control through deterrence on the compliance with information security policy. *Soft Computing*, *22*(20), 6765–6772.

D'Arcy, J., & Devaraj, S. (2012). Employee misuse of information technology resources: Testing a contemporary deterrence model. *Decision Sciences*, *43*(6), 1091–1124.

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., & Hovav, A. (2009). Does one size fit all? Examining the differential effects of IS security countermeasures. *Journal of Business Ethics*, *89*, 59–71.

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.

D'Arcy, J., & Lowry, P. B. (2019). Cognitive-affective drivers of employees' daily compliance with information security policies: A multilevel, longitudinal study. *Information Systems Journal*, *29*(1), 43–69.

Davis, M. S. (2006). Crimes mala in se: An equity-based definition. *Criminal Justice Policy Review*, *17*(3), 270–289.

Dhillon, G., Silva, L., & Backhouse, J. (2004). Computer crime at CEFORMA: A case study. *International Journal of Information Management*, *24*(6), 551–561.

Fan, J., & Zhang, P. (2011). Study on e-government information misuse based on general deterrence theory. *International Conference on Service Systems and Service Management*. https://doi.org/10.1109/ICSSSM.2011.5959454

Foth, M. (2016). Factors influencing the intention to comply with data protection regulations in hospitals: Based on gender differences in behaviour and deterrence. *European Journal of Information Systems*, *25*(2), 91–109.

Gibbs, J. P. (1975). *Crime, punishment, and deterrence*. Elsevier.

Gibbs, J. P. (1986). Deterrence Theory and Research. In G. Melton, L. Nader, & R. A. Dienstbier (Eds.), *Law as a Behavioral Instrument*. Lincoln: Univ. of Nebraska Press.

Gopal, R. D., & Sanders, G. L. (1997). Preventive and deterrent controls for software piracy. *Journal of Management Information Systems*, *13*(4), 29–47.

Green, S. P. (2016). The conceptual utility of malum prohibitum. *Canadian Philosophical Review*, *55*(1), 33–43.

Grover, V., & Lyytinen, K. (2015). New state of play in information systems research: The push to the edges. *MIS Quarterly*, *39*(2), 271–296.

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.

Hampton, J. (1986). *Hobbes and the Social Contract Tradition*. Cambridge, MA: Cambridge University Press.

Harrington, S. J. (1996). The effect of codes of ethics and personal denial of responsibility on computer abuse judgments and intentions. *MIS Quarterly*, *20*(3), 257–278.

Henle, C. a, & Blanchard, A. L. (2008). The interaction of work stressors and organizational sanctions on cyberloafing. *Journal of Managerial Issues*, *20*(3), 383–400. Retrieved from http://eserv.uum.edu.my/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=34354619&site=ehost-live&scope=site

Herath, T., & Rao, H. R. (2009a). Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. *Decision Support Systems*, *47*(2), 154–165. https://doi.org/10.1016/j.dss.2009.02.005

Herath, T., & Rao, H. R. (2009b). Protection motivation and deterrence: a framework for security policy compliance in organisations. *European Journal of Information Systems*, *18*(2), 106–125. https://doi.org/10.1057/ejis.2009.6

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 and People*, *31*(6), 1135–1162.

Higgins, G. E., Wilson, A. L., & Fell, B. D. (2005). An application of deterrence theory to software piracy. *Journal of Criminal Justice and Popular Culture*, *12*(3), 166–184.

Honderich, T. (2005). *Punishment: The Supposed Justification Revisited*. London: Pluto Press.

Hooper, V., & Blunt, C. (2019). Factors influencing the information security behaviour of IT employees. *Behaviour and Information Technology*, 1–13. https://doi.org/10.1080/0144929X.2019.1623322

Hovav, A., & D'Arcy, J. (2007). Deterring internal information systems misuse. *Communications of the ACM*, *50*(10), 113–117.

Hovav, A., & D'Arcy, J. (2012). Applying an extended model of deterrence across cultures: An investigation of information systems misuse in the U.S. and South Korea. *Information and Management*, *49*(2), 99–110. https://doi.org/10.1016/j.im.2011.12.005

Hu, Q., Xu, Z., Dinev, T., & Ling, H. (2011). Does deterrence work in reducing information security policy abuse by employees? *Communications of the ACM*, *54*(6), 54.

Hua, J., & Bapna, S. (2013). The economic impact of cyber terrorism. *Journal of Strategic Information Systems*, *22*(2), 175–186.

Ifinedo, P. (2014). Information systems security policy compliance: An empirical study of the effects of socialisation, influence, and cognition. *Information and Management*, *51*(1), 69–79. https://doi.org/10.1016/j.im.2013.10.001

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

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.

Joo, J., & Hovav, A. (2016). The influence of information security on the adoption of web-based integrated information systems.pdf. *Information Technology for Development*, *22*(1), 94–116.

Kankanhalli, A., Teo, H.-H., Tan, B. C. Y., & Wei, K.-K. (2003). An integrative study of information systems security effectiveness. *International Journal of Information Management*, *23*(2), 139–154. https://doi.org/10.1016/S0268-4012(02)00105-6

Karjalainen, M., Siponen, M., Puhakainen, P., & Sarker, S. (2013). One size does not fit all: Different cultures require different information systems security interventions. *PACIS*.

Khan, H. U., & Alshare, K. A. (2019). Violators versus non-violators of information security measures in organizations—A study of distinguishing factors. *Journal of Organizational Computing and Electronic Commerce*, *29*(1), 4–23.

Kim, B., Lee, D. Y., & Kim, B. (2019). Deterrent effects of punishment and training on insider security threats: a field experiment on phishing attacks. *Behaviour and Information Technology*, 1–20.

Kim, H. L., & Han, J. (2019). Do employees in a “good” company comply better with information security policy? A corporate social responsibility perspective. *Information Technology and People*, *32*(4), 858–875.

Knapp, K. J., Marshall, T. E., Rainer Jr, R. K., & Ford, F. N. (2007). Information security effectiveness: Conceptualization and validation of a theory. *International Journal of Information Security and Privacy*, *1*(2), 37–60. https://doi.org/10.4018/978-1-60566-196-4

Kohlberg, L. (1981). *Essays on Moral Development, Vol. I: The Philosophy of Moral Development*. San Francisco, CA: Harper & Row.

Kuo, K. M., Talley, P. C., Hung, M. C., & Chen, Y. L. (2017). A deterrence approach to regulate nurses’ compliance with electronic medical records privacy policy. *Journal of Medical Systems*, *41*(12), 1–10.

Lebek, B., Uffen, J., Neumann, M., Hohler, B., & Breitner, M. H. (2014). Information security awareness and behavior: A theory-based literature review. *Management Research Review*, *37*(12), 1049–1092.

Lee, J., & Lee, Y. (2002). A holistic model of computer abuse within organizations. *Information Management & Computer Security*, *10*(2/3), 57–63. https://doi.org/10.1108/09685220210424104

Lee, S. M., Lee, S.-G., & Yoo, S. (2004). An integrative model of computer abuse based on social control and general deterrence theories. *Information and Management*, *41*(6), 707–718.

Levin, A. M., Dato-on, M. C., & Manolis, C. (2007). Deterring illegal downloading: The effects of threat appeals, past behavior, subjective norms, and attributions of harm. *Journal of Consumer Behaviour*, *12*(4), 253–266.

Liao, Q., Luo, X., Gurung, A., & Li, L. (2009). Workplace management and employee misuse: Does punishment matter? *Journal of Computer Information Systems*, *50*(2), 49–59. Retrieved from http://search.ebscohost.com.ezproxy.liv.ac.uk/login.aspx?direct=true&db=iih&AN=48216614&site=eds-live&scope=site

Lowry, P. B., Dinev, T., & Willison, R. (2017). Why security and privacy research lies at the centre of the

information systems (IS) artefact: Proposing a bold research agenda. *European Journal of Information Systems*, *26*(6), 546–563. https://doi.org/10.1057/s41303-017-0066-x

Lowry, P. B., Posey, C., Bennett, R. (Becky) 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.

Luo, X. R., & Zhdanov, D. (2016). Special issue introduction: A comprehensive perspective on information systems security — technical advances and behavioral issues. *Decision Support Systems*, *92*, 1–2. https://doi.org/10.1016/j.dss.2016.10.003

Mahmood, A. M., Siponen, M., Straub, D., Rao, H. R., & Raghu, T. S. (2010). Moving toward black hat research in information systems security: An editorial introduction to the special issue. *MIS Quarterly*, *34*(3), 431–433.

Merhi, M. I., & Ahluwalia, P. (2019). Examining the impact of deterrence factors and norms on resistance to Information Systems Security. *Computers in Human Behavior*, *92*, 37–46.

Milne, S., Sheeran, P., & Orbell, S. (2000). Prediction and intervention in health-related behavior: A meta-analytic review of protection motivation theory. *Journal of Applied Social Psychology*, *30*(1), 106–143.

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

Nagin, D. S. (2013). Deterrence in the twenty-first century. *Crime and Justice*, *42*(1), 199–263.

Onwudiwe, I. D., Odo, J., & Emmanuel, O. (2004). Deterrence Theory. In B. Mary (Ed.), *Encyclopedia of Prisons and Correctional Facilities* (pp. 233–237). https://doi.org/doi: http://dx.doi.org/10.4135/9781412952514

Pahnila, S., Siponen, M., & Mahmood, A. (2007). Which factors explain employees' adherence to information security policies? An empirical study. *Pacific Asia Conference on Information Systems*.

Park, S. K., Kwak, K. T., & Lee, B. G. (2019). Policy compliance and deterrence mechanism in the sharing economy: Accommodation sharing in Korea. *Internet Research*, *29*(5), 1124–1148.

Paternoster, R. (2010). How much we really know about criminal deterrence? *The Journal of Criminal Law and Criminology*, *100*(3), 765–824.

Paternoster, R., Saltzman, L., Waldo, G., & Chiricos, T. (1983). Perceived risk and social control: Do sanctions really deter? *Law and Society Review*, *17*(3), 457–480.

Peace, A. G., Galletta, D. F., & Thong, J. Y. L. (2003). Software piracy in the workplace: A model and empirical test. *Journal of Management Information Systems*, *20*(1), 153–177.

Petrosino, A., Turpin-petrosino, C., & Finckenauer, J. O. (2000). Well-meaning programs can have harmful effects! Lessons from experiments of programs such as scared straight. *Crime & Delinquency*, *46*(3), 354–379.

Piquero, A. R., Paternoster, R., Pogarsky, G., & Loughran, T. (2011). Elaborating the individual difference component in deterrence theory. *Dx.Doi.Org*, *7*(1), 335–360.

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

Quinlan, M. (2010). Deterrence and deterrability. *Criminolgy*, *48*(2), 417–441.

Rajab, M., & Eydgahi, A. (2019). Evaluating the explanatory power of theoretical frameworks on intention to comply with information security policies in higher education. *Computers and Security*, *80*, 211–223.

Rowe, F. (2014). What literature review is not: Diversity, boundaries and recommendations. *European Journal of Information Systems*, *23*(3), 241–255.

Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J., Reuter, P., & Bushway, S. D. (1998). Preventing Crime: What Works, What Doesn't, What's Promising. In *(Report to the United States Congress). College Park: University of Maryland, Department of Criminology and Criminal Justice.* https://doi.org/10.1111/an.1986.27.3.11.5

Sinha, R. K., & Mandel, N. (2008). Preventing digital music piracy: The carrot or the stick? *Journal of Marketing*,

72(1), 1–15. https://doi.org/10.1509/jmkg.72.1.1

Siponen, M. (2006). A justification for software rights. *ACM SIGCAS Computers and Society*, *36*(3), 11–20.

Siponen, M., & Iivari, J. (2006). Six design theories for IS security policies and guidelines. *Journal of the Association for Information Systems*, *7*(7), 445–472.

Siponen, M., & Klaavuniemi, T. (2019). How and why 'theory' is often misunderstood in information systems literature. *Fortieth International Conference on Information Systems*, 1–16. Munich.

Siponen, M., & Klaavuniemi, T. (2020). Why is the hypothetico-deductive (H-D) method in information systems not an H-D method? *Information and Organization*, *30*(1), 100287.

Siponen, M., Pahnila, S., & Mahmood, A. (2007). Employees' adherence to information security policies: An empirical study. In R. Venter, H., Eloff, M., Labuschagne, L., Eloff, J., von Solms (Ed.), *IFIP International Federation for Information Processing* (Vol. 232, pp. 133–144). https://doi.org/10.1007/978-0-387-72367-9_12

Siponen, M., Pahnila, S., & Mahmood, M. A. (2010). Compliance with information security policies: An empirical investigation. *IEEE Computer*, *43*, 64–71. https://doi.org/10.1109/MC.2010.35

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

Siponen, M., & Vance, A. (2014). Guidelines for improving the contextual relevance of field surveys: the case of information security policy violations. *European Journal of Information Systems*, *23*, 289–305. https://doi.org/10.1057/ejis.2012.59

Siponen, M., Vance, A., & Willison, R. (2012). New insights into the problem of software piracy: The effects of neutralization, shame, and moral beliefs. *Information and Management*, *49*(7–8), 334–341.

Skinner, W. F., & Fream, A. M. (1997). A social learning theory analysis of computer crime among college students. *Journal of Research in Crime and Delinquency*, *34*(4), 495–518.

Sojer, M., Alexy, O., Kleinknecht, S., & Henkel, J. (2014). Understanding the drivers of unethical programming behavior: The inappropriate reuse of Internet-accessible code. *Journal of Management Information Systems*, *31*(3), 287–325.

Stanko, T. L., & Beckman, C. M. (2015). Watching you watching me: Boundary control and capturing attention in the context of ubiquitous technology use. *Academy of Management Journal*, *58*(3), 712–738. https://doi.org/10.5465/amj.2012.0911

Stoycheff, E., Liu, J., Xu, K., & Wibowo, K. (2019). Privacy and the Panopticon: Online mass surveillance's deterrence and chilling effects. *New Media and Society*, *21*(3), 602–619.

Straub, D. (1986). *Deterring computer abuse: the effectiveness of deterrent countermeasures in the computer security environment. PhD. Thesis*.

Straub, D. (1990). Effective IS security: An empirical study. *Information Systems Research*, *1*(3), 255–276.

Straub, D., & Nance, W. D. (1990). Discovering and disciplining computer abuse in organizations: A field study. *MIS Quarterly*, *14*(1), 45–60. https://doi.org/10.2307/249307

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

Templier, M., & Paré, G. (2015). A framework for guiding and evaluating literature reviews. *Communications of the Association for Information Systems*, *37*, 112–137.

Theoharidou, M., Kokolakis, S., Karyda, M., & Kiountouzis, E. (2005). The insider threat to information systems and the effectiveness of ISO17799. *Computers and Security*, *24*(6), 472–484. https://doi.org/10.1016/j.cose.2005.05.002

Thornberry, T. P. (1987). Toward an interactional theory of delinquency. *Criminolgy*, *25*(4), 863–892.

Toby, J. (1964). Is punishment necessary. *J. Crim. L. Criminology & Police Sci.*, *332*.

Trang, S., & Brendel, B. (2019). A meta-analysis of deterrence theory in information security policy compliance research. *Information Systems Frontiers*, 1265–1284.

Truex, D., Holmstrom, J., & Keil, M. (2006). Theorizing in information systems research: A reflexive analysis of the adaptation of theory in information systems research. *Journal of the Association for Information Systems*, *7*(12), 797–821.

Tsohou, A., Siponen, M., & Newman, M. (2020). How does information technology-based service degradation influence consumers' use of services? An information technology-based service degradation decision theory. *Journal of Information Technology*, *35*(1), 2–24.

Tunick, M. (1992). *Punishment: Theory and Practice*. Berkeley and Los Angeles, California: University of California Press.

Ugrin, J. C., & Pearson, J. M. (2013). The effects of sanctions and stigmas on cyberloafing. *Computers in Human Behavior*, *29*(3), 812–820.

Ugrin, J., Pearson, J., & Odom, M. (2008). Cyber-slacking: Self-control, prior behavior And the impact Of deterrence measures. *Review of Business Information Systems*, *12*(1), 75–88. https://doi.org/10.19030/rbis.v12i1.4399

Vance, A., & Siponen, M. (2012). IS security policy violations: A rational choice perspective. *Journal of Organizational and End User Computing*, *24*(1), 21–41.

von Hentig, H. (1938). Limits of deterrence. *Am. Inst. Crim. L. & Criminology*, *555*.

Webster, J., & Watson, R. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2), xiii–xxiii.

Weinstein, N. D., Lyon, J. E., Sandman, P. M., & Cuite, C. L. (1998). Experimental evidence for stages of health behavior change: the precaution adoption process model applied to home radon testing. *Health Psychology : Official Journal of the Division of Health Psychology, American Psychological Association*, *17*(5), 445–453.

Westcott, R., Ronan, K., Bambrick, H., & Taylor, M. (2017). Expanding protection motivation theory: investigating an application to animal owners and emergency responders in bushfire emergencies. *BMC Psychology*, *5*(13).

Williams, K. R., & Hawkins, R. (1986). Perceptual research on general deterrence: A critical review. *Law and Society Review*, *20*(4), 545–572.

Willison, R., Lowry, P. B., & Paternoster, R. (2018). A tale of two deterrents: Considering the role of absolute and restrictive deterrents in inspiring new directions in behavioral and organizational security. *Journal of the Association for Information Systems*, *19*(12), 1187–1216.

Willison, R., Warkentin, M., & Johnston, A. C. (2018). Examining employee computer abuse intentions: insights from justice, deterrence and neutralization perspectives. *Information Systems Journal*, *28*(2), 266–293.

Wittgenstein, L. (1953). *Philosophical Investigations*.

Wolfe, S. E., Higgins, G. E., & Marcum, C. D. (2008). Deterrence and digital piracy: A preliminary examination of the role of viruses. *Social Science Computer Review*, *26*(3), 317–333.

Workman, M., & Gathegi, J. (2007). Punishment and ethics deterrents: A study of insider security contravention. *Journal of the American Society for Information Science and Technology*, *58*(2), 212–222.

WoS. (2020). Web of Science Platform. Retrieved from Web of Science Group website: https://clarivate.com/webofsciencegroup/solutions/webofscience-platform/

Xue, Y., Liang, H., & Wu, L. (2011). Punishment, justice, and compliance in mandatory IT settings. *Information Systems Research*, *22*(2), 400–414. https://doi.org/10.1287/isre.1090.0266

Yang, C. G., & Lee, H. J. (2015). A study on the antecedents of healthcare information protection intention. *Information Systems Frontiers*, *18*(2), 253–263.

Yazdanmehr, A., & Wang, J. (2016). Employees' information security policy compliance: A norm activation perspective. *Decision Support Systems*, *92*, 36–46.

Zhang, L., Smith, W. W., & McDowell, W. C. (2009). Examining digital piracy: Self-control, punishment, and self-efficacy. *Information Resources Management Journal*, *22*(1), 24–44.

About the Authors

Mikko Siponen is a professor of information systems at the University of Jyväskylä. He has served as the vice dean for Research, the department head, vice head, and as the director of an IS security research center. His degrees include doctor of social sciences, majoring in applied philosophy; MSc in software engineering; and PhD in information systems. He has received several million euros in research funding from corporations and many other funding bodies. He is an invited member of The Finnish Academy of Science and Letters.

Wael Soliman is a post-doctoral research of Information Systems at the University of Jyväskylä. He earned his Dr. Sc. degree in Information Systems from Aalto University, School of Business, Finland. He received his M.Sc. degree in Information Systems from Hanken School of Economics, Finland. His research interests include IS security, IS usage and discontinuance, and issues related to IS research methods. His work is published in outlets such as Journal of the Association for Information Systems, Information and Management and European Journal of Innovation Management, Journal of Theoretical and Applied Electronic Commerce Research, and in the proceedings of leading IS conferences, such as the European Conference on Information Systems.

Anthony Vance is an associate professor and David Adamany Senior Research Fellow in the Department of Management Information Systems of the Fox School of Business at Temple University. He serves as the Director of the Center for Cybersecurity at the Fox School. He earned Ph.D. degrees in Information Systems from Georgia State University, USA; the University of Paris—Dauphine, France; and the University of Oulu, Finland. His previous experience includes working as a security consultant at Deloitte and as a research professor in the Information Systems Security Research Center at the University of Oulu. His research focuses on behavioral and neuroscience applications to information security. His work is published in outlets such as MIS Quarterly, Information Systems Research, Journal of Management Information Systems, Journal of the Association for Information Systems, European Journal of Information Systems, Journal of the American Society for Information Science and Technology, and Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI). He currently is an associate editor at MIS Quarterly and serves on the editorial board of Journal of the Association for Information Systems.

Appendix A

Review Procedure

To provide an accurate account on how deterrence theory has been used in the IS literature, we conducted a comprehensive literature review, which we describe in this section. Essentially, our review involves three core procedures: keyword search, auxiliary search, and extraction of (and synthesizing) key insights (Rowe, 2014; Templier & Paré, 2015; Webster & Watson, 2002). The main objective was to identify and include papers that apply DT in an IS context, to gain an understanding of the extent to which DT's assumptions have been treated in IS. The first stage was the ***keyword search procedure***. We decided to conduct the initial key word search on the Web of Science (WoS) because we wanted our search to be comprehensive and broad, and WoS met this objective since it covers thousands of scholarly journals, conference proceedings as well as hundreds of disciplines (WoS, 2020). Considering our stated focus on the application of DT in an IS context, we used the following search query: ('deterrence theory' AND ('information systems' OR 'information technology')) We set the timespan to 1945 to 2019, inclusive. This search query resulted in 147 hits.

Next, we extracted the abstracts of all 147 articles, and read them carefully to determine that the topic actually meets our two core criteria: a) it applies deterrence theory, and b) in an information systems/technology context. Based on this initial analysis, we divided the articles into three categories. The first category includes 35 articles that are clearly within the scope of our study, for example, the application of deterrence theory in the context of software piracy (Peace et al., 2003). The second category includes 70 articles that are clearly outside the scope of our study. Articles in this category fulfills only a single inclusion criterion, but not both. For example, the application of deterrence theory in a context irrelevant to IS/IT, e.g., drunk-driving (Albery & Guppy, 1995); or the study of an IS/IT context through a theoretical lens irrelevant to deterrence theory, e.g., the study of IT acceptance through the lens of emotions (Beaudry & Pinsonneault, 2010). The third category includes 42 articles where the abstract indicates that the article may or may not fulfill the inclusion criteria, thus requiring further inspection of the full text. Closer inspection of the 42 articles (in third category) led us to conclude that 11 articles fulfill the criteria, and that the remaining 31 articles will be excluded. This stage culminated in a total of 46 relevant studies.

In addition to the initial keyword search, we performed ***auxiliary search procedures*** with the objective of identifying relevant articles that might have been overlooked by WoS's search engine. Two auxiliary search procedures deemed helpful. First, with the help of backward search (Webster & Watson, 2002) where we scanned the bibliography list of the identified articles for potentially relevant references. Second, discussions with domain experts have brought to our attention further relevant articles that did not come up in the search procedures. Auxiliary search procedures helped further identify 40 relevant articles, thus, making the total number of articles applying DT in an IS context 84 articles.

Each paper was read carefully and the analysis involved two main activities: extraction and reflection (Templier & Paré, 2015). ***Extraction*** involved documenting important information such as source title, methodology, violation context, theoretical base, deterrence elements, and operationalization of deterrents, and most importantly, the extent to which the eight assumptions have been discussed or acknowledged. ***Synthesis***, on the other hand, involved taking notes on the key findings, interesting insights, original authors' opinions and quotes, as well as, our own personal assessment of the article in general. The main aim here is to deliver 'a global representation' of the deterrence-oriented IS literature (Rowe, 2014).

Appendix B

Table B. The Deterrence Hypothesis Organized by Context and Level of Criminality

Context	Criminality	Deterrence Hypothesis in Specific Context
Organizational	Criminal (or partly criminal)	Supported in 5 studies: • Computer abuse/misuse (D'Arcy & Devaraj, 2012; Straub, 1990; Straub & Nance, 1990). • Software piracy (Peace et al., 2003). • Coping with system risk (Straub & Welke, 1998).
		Partially supported in 9 studies: • Computer abuse/crime (Dhillon et al., 2004; Harrington, 1996; Lee et al., 2004; Nicho & Kamoun, 2014; Willison et al., 2018). • Information misuse (D'Arcy & Hovav, 2009; D'Arcy et al., 2009; Fan & Zhang, 2011; Hovav & D'Arcy, 2012)
		Not supported in 1 study: • Illegal policy violation (Hu, Xu, Dinev, & Ling, 2011)
	Noncriminal	Supported in 18 studies: • Policy violation (Alshare et al., 2018; Barlow et al., 2013; Herath et al., 2018; Johnston et al., 2016; Khan & AlShare, 2019; Workman & Gathegi, 2007). • Policy resistance (Merhi & Ahluwalia, 2019). • Policy compliance (Choi & Song, 2018; Connolly et al., 2017; Kim & Han, 2019; Pahnila et al., 2007; Siponen et al., 2007; Siponen et al., 2010; Yazdanmehr & Wang, 2016). • Employees' response to phishing attacks (B. Kim et al., 2019). • Personal Internet use at work (Henle & Blanchard, 2008). • Cyber-slacking (J. Ugrin et al., 2008). • IS security effectiveness (Kankanhalli et al., 2003).
		Partially supported in 13 studies: • Policy violation (Cheng et al., 2013; Kuo et al., 2017). • Policy compliance (Bulgurcu et al., 2010; X. Chen et al., 2018; Y. Chen et al., 2012; Foth, 2016; Herath & Rao, 2009a, 2009b). • Security effectiveness (Knapp et al., 2007). • Healthcare information protection (Yang & Lee, 2015). • Personal Internet use at work (Liao et al., 2009; Ugrin & Pearson, 2013). • Re-use of programming code (Sojer et al., 2014).
		Not supported in 11 studies: • Policy violation (Guo et al., 2011; Lowry et al., 2015; Siponen & Vance, 2010; Vance & Siponen, 2012). • Policy compliance (Hooper & Blunt, 2019; Ifinedo, 2014; Johnston et al., 2015; Moody et

		al., 2018; Rajab & Eydgahi, 2019; Xue et al., 2011). • Technology use (Joo & Hovav, 2016).
Non-organizational	Criminal (or partly criminal)	Supported in 2 studies: • Software piracy (Gopal & Sanders, 1997). • Online (il)legal behavior (Stoycheff et al., 2019).
		Partially supported in 5 studies: • Computer crime (Skinner & Fream, 1997). • Digital and software piracy (Higgins et al., 2005; Levin et al., 2007; Sinha & Mandel, 2008; Zhang et al., 2009).
		Not supported in 2 studies: • Digital/software piracy (Siponen et al., 2012; Wolfe et al., 2008).
	Noncriminal	Supported in 1 study: • Sharing economy policy compliance (Park et al., 2019).

Appendix C

Table C. Recognized Assumptions and Empirical Application

Article	Notes on the acknowledged assumptions
(D'Arcy & Herath, 2011)	This article acknowledges 4 assumptions related to the act (#1 and #2), the actor (#5) and the punishment (#6). **#1:** Authors acknowledge that "the perceived cost portion of the rational decision process may weigh heavier for those less common behaviors that are more disruptive of IS security, such as intentional abuse of IS resources, because employees likely perceive stronger negative consequences (i.e., stiffer penalties) for engaging in such activity. This is in line with findings that the threat of punishment is more important in deterring serious criminal activity than in deterring minor offenses" (p. 653). **#2:** Authors argue that one possible explanation for the weak influence of sanction on 'positive outcome variables' can be attributed to 'social control theory' which "argues that attachment to conventional norms, and not so much the threat of sanctions, is really what drives conformity (Cao, 2004)" (p. 652). **#5:** Although this issue is not explicitly discussed, the authors note that "self vs other-referenced perceived sanction measures" as a potential source of the discrepancy in DT studies' findings. But then they use a different argument in terms of measurement items (asking respondents about organizational punishments in general vs. in specific scenarios) assuming that respondents will consider scenarios more "self-referenced". **#6:** Authors argue one of the possible reasons for conflicting findings is using different measures for severity of sanction, and highlight that fixed measures (e.g., ranging from 'no fine' ... '5-years jail') "assume that a particular punishment has the same meaning for all people ... The fallacy of this approach is that what is felt as extremely costly for one individual (e.g., $1000 fine) may be considered insignificant for another" (p.651).
(Ugrin & Pearson, 2013)	This article acknowledges 3 assumptions related to the act (#2), the actor (#5) and the punishment (#8). **#2:** Authors argue that "abusiveness perception" moderates the effect of sanction on behavior. This resonates with the notion that sanctions are not effective in situations where behavior is personally/socially accepted. **#5:** Authors argue that "It is reasonable to conclude from our findings that AUPs and Internet monitoring will be relatively ineffective at reducing behaviors like personal emailing and social networking unless employees know about of others who have been caught and punished severely" (p. 818). **#8:** While the findings suggest that severe punishment is necessary to reduce cyberloafing behaviors like personal emailing and social networking, they also acknowledge its negative consequences (calling the whole quandary a "catch-22"): "we do not account for negative affect that can be a result of the use of deterrence mechanisms and detection and research has shown that simply introducing these types of deterrence methods can create employee strife" (p. 818). **Empirical application:** Assumption #5 is invoked (post-hoc) to explain the nonsignificant impact of sanction on violation.
(Willison, Lowry, et al., 2018)	This article acknowledges 3 assumptions related to the act (#1 and #3) and the actor (#5). **#1:** Authors explain the inconsistent DT-based results by noting that: 1. "DT was designed not to motivate good behavior or to explain how to thwart noncriminal noncompliance, but to explain and predict how to thwart criminal or antisocial behavior" (p. 1189) .. 2. "DT has been erroneously contextualized in most ISec studies because few of them have focused on criminal behavior, and the omission could be the reason for the disparate findings in the ISec DT literature" (p. 1189) **#3:** Author state explicitly the importance of knowledge of offender's knowledge of law and applicable organizational policies. **#5:** Authors refer to "developmental criminology", which looks at crimes from a stage-like perspective, with different decision points (beginning, continuation, cessation). Also, in their discussion of the distinction between absolute and restrictive deterrence, the authors acknowledge the distinction between initial involvement (i.e., beginning a criminal career), and decisions made by repeating offenders.
(Straub, 1990)	This article acknowledges 2 assumptions related to the act (#3) and the actor (#5). **#3:** In the discussion section, Straub recommends that in order to improve security, managers "should establish policies regarding proper and improper use of the information system". He also notes that "a clearly defined set of policies is the precondition to implementing all the effective deterrents". He also proposes that "IS security officers should inform and educate users about acceptable system use" (p. 272). **#5:** Author highlights that GDT focuses on "sanctions against committing a deviant act and the effect of these sanctions on deterring others from committing criminal acts." (p. 258).

(Lee & Lee, 2002)	This article acknowledges 2 assumptions related only to the act (#2 and #3). **#2:** Authors refer to "social bond theory" as giving additional explanation by accounting for situations in which deterrents do not work, e.g., when the crime is perceived positively among peers; and when deviant behavior is deterred not because of certainty & severity of sanction but because of "a sense of fairness". **#3:** Authors implicitly acknowledge the knowledge assumption by highlighting that "belief" component in social bond theory (whether a person believes behavior is illegal or not) is critical to its enactment.
(Lee et al., 2004)	This article acknowledges 2 assumptions related only to the act (#2 and #3). **#2:** Building on the social control theory, the authors highlight the importance social factors in preventing computer abuse. They found that mainly "Norms" and "Involvement" had a significant and negative impact on (the induction control part of) computer abuse intentions. **#3:** The study highlights the deterring effect of "security awareness" in general, however, there was no significant impact on (the self-defense part of) computer abuse intentions.
(Theoharidou et al., 2005)	This article acknowledges 2 assumptions related only to the act (#1 and #3). **#1:** Authors highlight that ISO17799 does not emphasize the use of severe sanctions. Also, they implicitly question the relevance of sanctions in insider threat situations where "people do not always make rational calculated decisions [and that] their actions are often dictated by anger, frustration or despair" (p. 480). **#3:** Authors implicitly addressed by highlighting the "deterrence feedback".
(Siponen et al., 2007)	This article acknowledges 2 assumptions related only to the act (#2 and #3). **#2:** Authors emphasize that social pressure (disapproval) is important for ensuring ISP compliance. **#3:** Authors emphasize that practitioners need to state the sanctions for ISP non-compliance in a "visible manner".
(Knapp et al., 2007)	This article acknowledges 2 assumptions related to the act (#2) and the punishment (#8). **#2:** Security awareness and training emerged as a key predictor for security effectiveness in the GTM phase. This was also confirmed during the SEM phase **#8:** Closely related here is "the dilemma of the supervisor", which is described by Strickland (1958) as "occuring when the excessive use of surveillance, monitoring, and authority leads to management's distrust of employees and the perception of an increased need for more surveillance and control. Because managers see employees as motivated by the controls in place, managers develop a jaundiced view of their people (Ghoshal, 2005, p. 85). Too much surveillance and monitoring of employee activities to enforce policy compliance is then perceived by employees as overly controlling, which may damage employee self-perception, deteriorate trust, and decrease intrinsic motivation (Ghoshal, 2005)" (p. 54).
(Siponen et al., 2010)	This article acknowledges 2 assumptions related only to the act (#2 and #3). **#2:** Authors highlight the importance of social pressure as a deterrence. **#3:** Authors highlight that awareness is key. They recommend that managers ensure that information security staff make employees aware of security threats, their severity, and celerity.
(Guo et al., 2011)	This article acknowledges 1 assumption related to the act (#1). **#1:** Pointing to the inconclusive evidence regarding the effect of deterrence on IS misuse the authors argue "deterrence models may help explain why users comply with computer use or security rules (by not engaging in NMSVs), but not why they break these rules or engage in NMSVs" (p. 208-209). They also Provides support that sanctions have no significant effect in non-malicious IS violations contexts. **Empirical application:** Assumption #1 is incorporated in study design by focusing specifically on non-malicious ISP violations. Empirical evidence suggests that sanctions have no significant effect in non-malicious ISP violations.
(D'Arcy & Devaraj, 2012)	This article acknowledges 2 assumptions related to the act (#1) and the actor (#5). **#1:** Although not directly related, the authors acknowledge that formal sanctions is more relevant to technology misuse and other white-collar offenses that are more calculated and instrumental, compared to spontaneous and emotion-driven offences where formal sanctions might not explain such behaviors. **#5:** Authors argue that scenario is a suitable approach due to the temporal nature of the deterrence process. "Deterrence researchers have raised the issue of experiential effects in that people who previously committed illicit acts and got away with them may have lower perceptions of FS in the present (Williams & Hawkins, 1986)"

(D'Arcy and Hovav, 2009)	This article acknowledges 2 assumptions related to the act (#3) and the actor (#6). **#3:** highlight the "knowledge" aspect, i.e., their model incorporates "users' awareness of security countermeasures" as antecedents to forming "sanction perceptions". **#6:** Authors highlight the differential deterrence hypothesis (Mann et al. 2003), "which posits that the impact of formal sanctions is not uniform across all persons due to individual and situational differences (Mann et al., 2003). Such differences are thought to influence sanction perceptions, which in turn have a direct effect on behavior (Tittle, 1980)" (p. 61).
(Higgins et al., 2005)	This article acknowledges 2 assumption related to the actor (#5) and the punishment (#6). **#5:** Authors note that deterrence theory aims to reduce deviant and criminal behavior "by identifying, discovering, and properly penalizing individuals who are currently performing criminal behaviors. On the other hand, the criminal justice system threatens individuals not currently performing criminal behaviors with legal sanctions (i.e., the certainty and severity of punishment) in order to control crime (p. 169). **#6:** Authors acknowledge this assumption by noting the "perceptual nature of deterrence theory" (p. 173) **Empirical application:** Assumption #6 is applied by involving the study participants in defining the link between objective and perceptive sanctions (p. 173).
(Lowry et al., 2017)	This article acknowledges 2 assumptions related only to the act (#1 and #3). **#1:** Authors acknowledge that DT was originally developed for criminal contexts, and applying it in non-criminal contexts requires thoughtful consideration of its assumptions and logic, thus producing a "different version of DT" **#3:** The authors advise future research in their DT recontextualization efforts to clarify "what kinds of non-legal formal and informal sanctions work best with various kinds of security/privacy behaviours, and ... why." (p 553)
(Moody et al., 2018)	This article acknowledges 2 assumptions related to the act (#1) and the actor (#5). **#1:** Authors found no significant impact of sanction on the three scenarios (password sharing, locking computer, and USB practices) and thus argue that DT might not be relevant in those contexts. **#5:** Authors acknowledge general deterrent effect and use it to explain why DT did not have significant results. They argue that a potential explanation for the non-significant results is "the lack of the deterrence effect (Gibbs 1975)—namely, that sanctions are not effective if there are no examples of people who have been caught" (p. 308). **Empirical application:** Assumption #1 is invoked to explain (post-hoc) the nonsignificant impact of sanction ISP compliance. Assumption #5 is invoked to explain (post-hoc) the nonsignificant results–"namely, that sanctions are not effective if there are no examples of people who have been caught" (p. 308).
(Stoycheff et al., 2019)	This article acknowledges 2 assumptions related to the actor (#5) and the punishment (#8). **#5:** "As people encounter new information— like others' consequences for similar behaviors or new practices and policies—they incorporate it into their cost–benefit assessments and update their probabilities accordingly (Anwar and Loughran, 2011). Thus, an individual's decision is a combination of his or her past experiences and the acquisition of new information" (p. 605). **#8:** Authors point to the "chilling effect" whereby surveillance has a deterrent effect on illegal behavior (piracy) but also transcends to legal behavior as well, such political engagement behavior. **Empirical application:** Assumption #8 is invoked to show illustrate the "chilling effect" as an undesired consequences of sanctions.
(Park et al., 2019)	This article acknowledges 2 assumptions related only to the actor (#5 and #6). **#5:** Authors point to the developmental nature of deterrence in terms of how deterrent effect is learned: "In terms of the deterrence theory, it has been suggested that the social environment affects policy compliance and non-compliance as a norm. A higher number of non-compliant peers indicates that individuals are more likely to recognize that their non-compliance behavior is acceptable and thus feel less guilty." **#6:** Authors acknowledge that "This theory [DT] has also been used to propose that deterrence is influenced by individuals' personal perceptions of punishment rather than the objective existence of the punishment itself (Williams and Hawkins, 1986) " (p. 1127). **Empirical application:** Assumption #5 is invoked to explain (post-hoc) the nonsignificant impact of sanction ISP (non)compliance.
(Kim et al., 2019)	This article acknowledges 2 assumptions related to the actor (#5) and the punishment (#8). **#5:** Authors acknowledge the distinction between general and specific deterrence, and acknowledge the learning nature of deterrent effect: "... without actual sanction experience, one can learn about the disadvantages resulting from anti-policy behaviour by observing rule breakers being punished or putting oneself in their shoes ...

	This fear of penalties decreases one's motivation to disobey security regulations and, thus, creates the general deterrent effect from perceived punishment" (p. 3). **#8:** Authors acknowledge that severe sanctions may backfire: "In general, the sanctions expected for failing to avoid a phishing attack are lenient rather than harsh (Wombat Security 2018). In fact, the sanction should not be too severe because it can cause adverse effects" (p. 6).
(Wolfe et al., 2008)	This article acknowledges 2 assumptions related to the actor (#5) and the punishment (#6). **#5:** Authors makes the distinction between general vs. specific and defines them properly. **#6:** Author acknowledge the perceptual nature of sanction via including the participants in defining the "bad things" that might occur as punishment. **Empirical application:** Assumption #6 is applied by involving the study participants in defining the link between objective and perceptive sanctions (p. 324).
(Peace et al., 2003)	This article acknowledges 1 assumption related to the act (#3). **#3:** Authors highlight the importance of clearly communicating what is punishable: "Simply having the rules on the books will do little to create change, if the rules are not enforced. Punishments should be clearly defined and communicated to the relevant audience" (p. 168).
(Balozian & Leidner, 2017)	This article acknowledges 1 assumption related to the act (#3). **#3:** Authors acknowledge that "Since the sanctions are in and of themselves insufficient for enforcing compliance, they must be communicated to IS users during security training, and the employees must be well informed about the penalties of breaching security" (p. 18).
(Harrington, 1996)	This article acknowledges 1 assumption related to the act (#3). **#3:** Author notes that "codes [of ethics] are believed to deter computer abuse because they keep employees abreast of laws and regulations and clearly define unacceptable or illegal conduct" (p. 259)
(Skinner & Fream, 1997)	This article acknowledges 1 assumption related to the actor (#5). **#5:** The authors argue that one reason why deterrents had no significant effect in their study is that, "there is neither a general nor specific deterrent effect that serves as a differential reinforcement/punishment for computer crime" (p. 513). Especially so when considering that only 7.3% of respondents were caught (specific), only 13.3% had friends who were caught (general) after illegal computer activity. **Empirical application:** Assumption #5 is invoked to explain (post-hoc) the non-significant impact of deterrence on violation.
(Straub & Welke, 1998)	This article acknowledges 1 assumption related to the act (#3). **#3:** They Highlight the importance of security policies and awareness programmes, which convey, among others, knowledge about sanctions. "In essence, potent security awareness training stresses the two central tenets of general deterrence theory: certainty of sanctioning and severity of sanctioning" (p. 445). In fact, security awareness programmes is one of the three risk coping elements the study proposes.
(Dhillon et al., 2004)	This article acknowledges 1 assumption related to the punishment (#8). **#8:** Authors point to a research stream that links workplace conditions to crime, e.g., because "operational level workers are often subjected to a high level of surveillance they often end up becoming disgruntled. As a consequence they have a greater probability of becoming involved in a crime" (p. 552). Furthermore, the findings show that "excessive controls following a criminal act disrupted the organisational activities" (p. 554), e.g., due to excessive controls research staff felt alienated, and as a consequence most research projects failed to finish on time and ended up being over budget (p. 557). **Empirical application:** Assumption #8 is invoked to portray the negative consequences of punishment.
(Workman & Gathegi, 2007)	This article acknowledges 1 assumption related to the act (#2). **#2:** Authors argue that a person's attitude towards law is largely shaped by person's legal socialization through: a) societal imposition, b) state-law constraints, and c) human interaction (reciprocal interaction). They also argue that "socially conscious" people would follow rules out of a sense of social responsibility and social conformity more than out of fear of punishment.
(Hovav & D'Arcy, 2007)	This article acknowledges 1 assumption related to the act (#3). **#3:** Authors emphasize users' awareness and educating users of issues like appropriate use, what constitutes legitimate use, etc.
(Pahnila et al., 2007)	This article acknowledges 1 assumption related to the act (#3). **#3:** Authors emphasize that practitioners need to state the sanctions for ISP non-compliance in a "visible manner".

(Levin et al., 2007)	This article acknowledges 1 assumption related to the actor (#6). **#6:** This study utilized an interesting approach to define weak vs. strong threats. They offered participants 9 types of punishments as a consequence of illegal music downloading. Based on the means of the results, they created three-tier threat manipulation (weak, moderate, strong). **Empirical application:** Assumption #6 is applied by involving the study participants in defining the link between objective and perceptive sanctions (p. 116).
(Sinha & Mandel, 2008)	This article acknowledges 1 assumption related to the punishment (#8). **#8:** The results from this suggests that negative incentives are a strong deterrent for certain consumers but can actually increase piracy tendencies for others: "for consumers with high levels of optimum stimulation (and, thus, higher tolerance for risk), increasing the perceived risk might actually backfire by slightly increasing their likelihood to pirate" (p. 12) **Empirical application:** Assumption #8 is invoked to illustrate empirically (post-hoc) the negative consequences of sanctions for actors with high risk tolerance.
(Zhang et al., 2009)	This article acknowledges 1 assumption related to the act (#1). **#1:** Authors note that digital piracy is acceptable to many people because it is considered “a soft crime which seems harmless, with no violence involved and, therefore, of no real danger to anyone” (p. 25).
D'Arcy et al.	This article acknowledges 1 assumption related to the act (#3). **#3:** highlight the "knowledge" aspect, i.e., their model incorporates "users' awareness of security countermeasures" as antecedents to forming "sanction perceptions".
(Herath and Rao 2009)	This article acknowledges 1 assumption related to the punishment (#8). **#8:** Unexpectedly, severity of penalty was found to have a significant and negative impact on intentions to comply with security policy. Authors point to Oliver's (1980) explanation that "negative incentives are effective in motivating unanimous cooperation, but their use is often uneven, cyclical and may generate hostilities which disrupt that cooperation they enforce" (p. 162). **Empirical application:** Assumption #8 is invoked (post-hoc) to explain the unexpected empirical findings suggesting the negative consequences of sanctions on compliance intentions.
(Bulgurcu et al., 2010)	This article acknowledges 1 assumption related to the act (#3). **#3:** Authors argue that "the objective of creating information security awareness is to make employees cognizant of risks related to information security and to educate them about their roles and responsibilities concerning those risks" (p. 528). The study found that ISA, which is formed by general ISA and ISP awareness, influences an employee’s attitude to comply with the ISP directly, as well as indirectly, through the employee’s compliance-related outcome beliefs.
(Xue et al., 2010)	This article acknowledges 1 assumption related to the actor (#5). **#5:** Although the authors use DT in a trivial manner and make no reference to generic/specific discussion, they captured actual punishment for the surveyed employees. Related to this discussion is the finding that actual punishment is negatively affects perceived justice of punishment, while at the same time the latter positively affects compliance intentions. The authors here point out that observers develop a perception of risk regarding the inappropriate IT usage behavior that led to the negative consequences.
(Hovav & D'Arcy, 2012)	This article acknowledges 1 assumption related to the actor (#6). **#6:** Authors argue and show empirically that people from different cultures may respond to sanctions differently.
(Chen et al., 2012)	This article acknowledges 1 assumption related to the punishment (#8). **#8:** Commentary from the study subjects suggest that not only excessive sanctions may backfire; but also that emphasis on sanctions only create an unhealthy atmosphere. Subjects reported that "they dislike the “unmotivated atmosphere” caused by the “pure” sanction enforcement policy of their organization" (p. 177).
(Hua and Bapna, 2013)	This article acknowledges 1 assumption related to the actor (#6). **#6:** Authors note that "It is well known that terrorists are willing to sacrifice their lives to inflict damage to the target so more efforts should be spent to deter cyber terrorists." (p. 183)
(Chen and Li 2014)	This article acknowledges 1 assumption related to the punishment (#8). **#8:** Though not well-articulated, the authors provide an explanation why excessive punishment may have a negative impact on compliance in the long run (i.e., backfire) in p.6.
(Lowry eta al., 2015)	This article acknowledges 1 assumption related to the punishment (#8). **#8:** Although authors do not specifically study the backfire effect of sanctions, based on Reactance theory, they refer to the backfire effect of increasing controls in a work environment.

(Burns et al., 2017)	This article acknowledges 1 assumption related to the actor (#5). **#5:** Authors acknowledge the distinction between generic and specific, but adopts the generic view. Quotes Yu (1994): "The deterrence theory focuses on three dimensions: celerity, certainty, and severity; that is, fast, certain, and harsh punishment discourages those punished from violating again (specific deterrence) and dissuades the general public, which learns about the punishment from committing crimes (general deterrence)" (p. 355).
(Kuo et al., 2017)	This article acknowledges 1 assumption related to the punishment (#8). **#8:** Considering the insignificant impact of sanction severity on nurses violation intention, the authors argued that a plausible explanation "may be that rigorous punishments may weaken the nursing staff's trust or loyalty toward hospitals and, thus, exerting a counterproductive influence on their compliance intention of the stated privacy policy" (p. 8).
(Chen et al., 2018)	This article acknowledges 1 assumption related to the act (#1). **#1:** The authors found no significant moderating effect of sanction on compliance intentions. Thus they questioned the direct impact of sanction on ISP compliance, noting that "GDT theory was developed in the context of criminality of a higher consequence such as incarceration, loss of assets, job, etc. In most circumstances, however, violation of security policy, does not rise to the level that was originally stated in the GDT theory development. Therefore, formal sanctions may not be as serious concern to employees in the context of ISP compliance as it is in a criminal context ... For example, employees may feel that the sanction severity may be higher if being caught of leaking confidential data as compared to choosing a weak password. In other words, formal sanction severity may vary depending on the violation types." (p. 1056). Empirical application: Assumption #1 is invoked to explain (post-hoc) the nonsignificant impact of sanction on ISP compliance.
(Choi and Song, 2018)	This article acknowledges 1 assumption related to the actor (#5). **#5:** Authors acknowledge the distinction between specific and general deterrence, although this does not seem to have any impact on how they conducted the study.
(Trang and Brendel, 2019)	This article acknowledges 1 assumption related to the act (#1). **#1:** Authors note that "deterrence theory originally was conceptualized in criminology to predict criminal behavior (Gibbs 1975). Studies typically have been conducted in contexts involving theft, vandalism, bribery, white- collar crimes, or price rigging" (p. 1268)
(Merhi & Ahluwalia, 2019)	This article acknowledges 1 assumption related to the actor (#5). **#5:** Authors argue that punishments should be made public to deter others. They note: "Stories of employees who received organizational punishment because of non-compliance can be spread-using blogs, newsletters, and e-mails, so others become aware of the consequences of non-compliance. This makes employees know that ISS policies is the right thing to do and others around are applying these policies and not resisting them." (p. 44)
(Rajab & Eydgahi, 2019)	This article acknowledges 1 assumption related to the punishment (#8). **#8:** opposite effect "The findings of this study suggest as the probability of detection goes higher, the intentions of employees' compliance with information security goes lower" (p. 221). **Empirical application:** Assumption #8 is invoked to illustrate (post-hoc) the negative consequences of sanction on compliance intentions.

Appendix D

Table D. Summary of Future Research Directions in IS

DT assumptions warranting future IS research	Future research directions	
	General recommendations	**Example of specific research direction**
#1. Criminality of the (punishable) act	• Examine the extent to which DT may be applicable to different organizational cultures and contexts. • Develop a new IS-specific theory to explain deterrence of non-malicious violations.	• Explore the reasoning (e.g. thought processes) of employees that deem certain non-malicious security violations acceptable. • Explore how such thought process (or the applicability of non-malicious violations) could be linked to the security criticality of business. • Examine whether different organizational cultures towards information security (e.g. bank and military versus university) explain the success of applying DT to non-malicious violations? For example, are the information security policies and guidelines viewed deontologically, as Kant's categorical imperatives (i.e. to-the-letter and absolutely), or are they regarded in a more flexible manner as a supererogative matter? • Develop measures for testing how strong the information security culture is in an organization.
#2. Social contract underlying the (punishable) act	• Investigate how to create social contracts for non-malicious ISP violations in which harm is not obvious to the people involved. • Explore the possible interplay between social contract and the business criticality of the organization.	• Explore whether the study participants believe in (and are committed to) their organization's ISP and sanctions or view them as unjustified and/or disruptive. • Examine the role of the strangeness of organizational information security cultures and how it links to social contract (e.g. social contract behind non-malicious violations in military or banks versus university). • Perform an action research study aimed at ascertaining how different educational strategies and argumentations can be used to discourage unwanted information security behaviors.
#3. Knowledge of the (punishable) act	• Make sure that actors in the studied context are aware of the ISP and are sufficiently knowledgeable of what is punishable or sanctionable.	• Develop tests that go beyond measuring employees' awareness of the existence of information security policies and measure their awareness of the actual content of these policies. • Explore assumptions regarding what makes a certain work environment 'critical', a certain violation 'malicious', and the cognitive processes deeming such violations 'sanctionable'.
#4. Actors are self-interested	• Examine to what extent people are self-interested. • Explore broader conceptions of self-interest, for example, greedy self-interest vs. enlightened self-interest, or the belief that benefiting the group will benefit the individuals	• Explore the extent to which certain IS security behaviors are linked to self-interest in the traditional narrow sense or in a more enlightened sense. • Develop and test instruments designed to measure individuals' levels of moral development in the context of IS security. • For example, make simple tests for each Kohlberg stage of

	belonging to that group. • Explore whether sanctions deter only self-interested people.	cognitive moral development.
#5. Actors learn to abstain via general or specific deterrence	• Application of generic and specific separately in different IS security contexts. • Examine which types of sanctions are more relevant or effective in IS security contexts. • Explore how fear from sanction regarding ISP violations may be learned over time.	• Develop instrumentation that makes a clear distinction between general and specific deterrents. The following is an example: o 'I have seen that others have received sanction X for doing Y' (general deterrence) o 'I have received sanctions S for X' (specific deterrence) • Explore whether general deterrence is based on seeing, observing, or hearing about sanctions from others. • Explore the significance of others' closeness, for example, knowing that someone somewhere received a sanction for committing an ISP violation vs. a colleague sharing the same office space. • Develop process or stage theories to explain how deterrent effects are learned over time.
#6. Punishment experience is subjective	• Explore the link between properties of prescribed sanctions and how they are perceived. • Explore what kind of sanctions are considered severe by actors in the studied context and explore their impact as deterrents.	• Develop instrumentation that better captures the subjective nature of sanction severity (see, e.g. Levin et al., 2007). • Interviews of people as to what is severe (enough) for them for different ISP violations. • Scenario-based message manipulations (e.g. Jasso, 2006) that present a range of specific sanctions. • Field experiments in companies where employees are menaced by certain specified sanctions; it is examined how people react to these by using different measures.
#7. Punishment for one act may deter another	• Explore whether the deterrent effect is transferable in IS security contexts. This research direction should take into consideration the distinction made earlier regarding the source of the deterrent effect.	• Examine the extent to which having personally experienced a sanction S for doing a violation X is transferable to other IS security violations (e.g., Y and Z). • Examine the extent to which having observed someone receiving a sanction S for doing a violation X is transferable to other IS security violations (e.g., Y and Z).
8. Punishment may backfire	• Explore the various facets in which backfiring may occur (e.g. hampered loyalty, increased negligence in other areas, revenge, etc.). • Explore phenomena beyond deterrence, for example, behaviors that emerge post-sanctions.	• Investigate the potential negative effects caused by unjustified sanctions, such as decreases in organizational commitment and security attitudes. • Develop instrumentation to determine the level of employees' moral development. • Examine how different moral reasoning levels may correlate with different ISP violations.

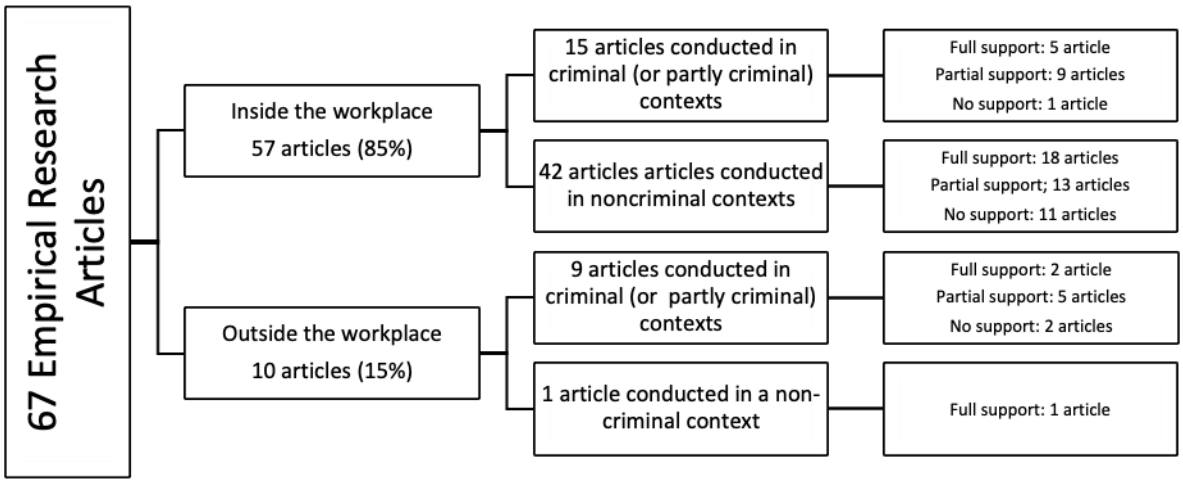

Figure 1. Overview of Empirical Evidence of the Deterrence Hypothesis in IS

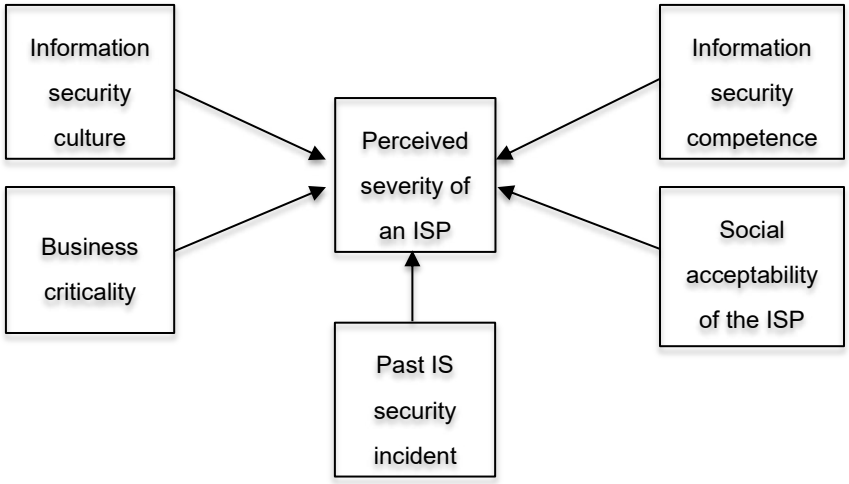

Figure 2. Possible Antecedents of the Perceived Severity of an ISP Violation.

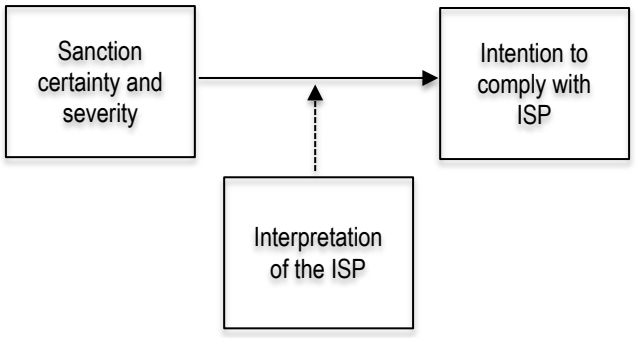

Figure 3.1. ISP Interpretation as Moderator

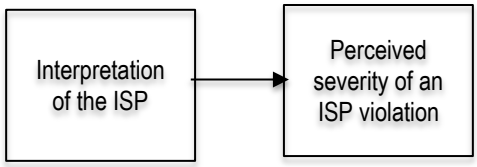

Figure 3.2. ISP Interpretation as Independent Variable

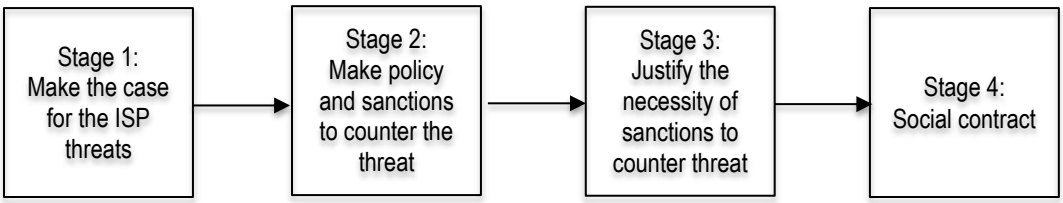

Figure 4. Simplistic Illustration of the Stages for Creating a Social Contract

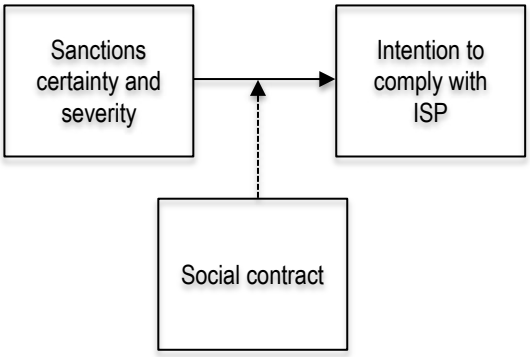

Figure 5.1. Social Contract as a Moderator

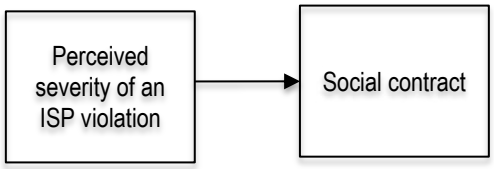

Figure 5.2. Social Contract as a Dependent Variable

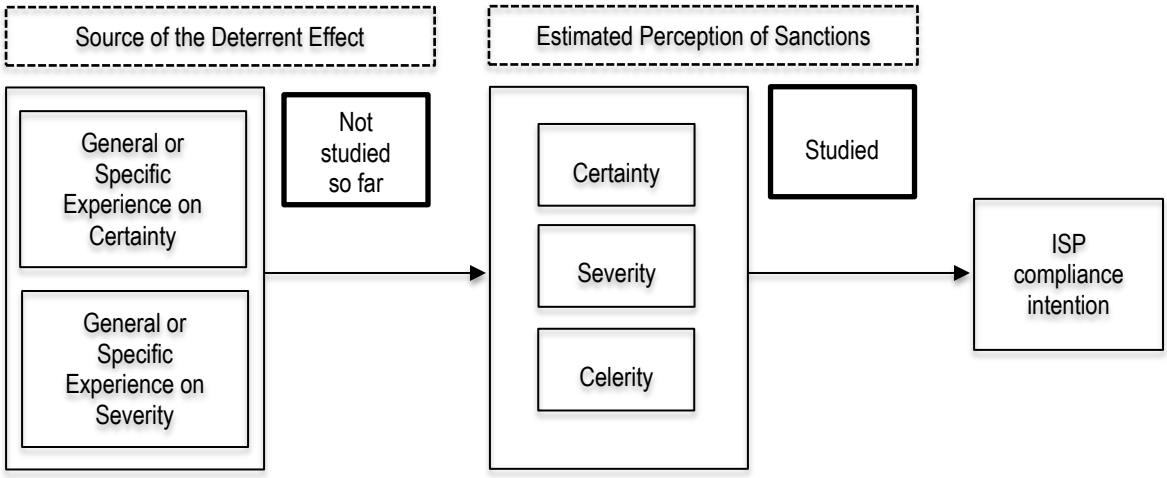

Figure 6. General and Specific Deterrent Effects and their Relationships with Estimated Future Perceptions of Sanctions

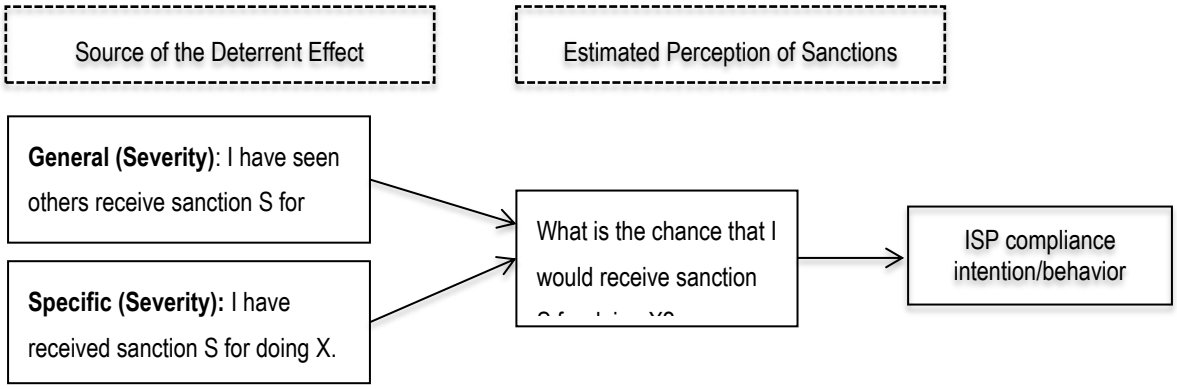

Figure 7. General and Specific Deterrent Effects and their Relationships to Estimated Future Perception of Sanctions; Example of Sanction Severity

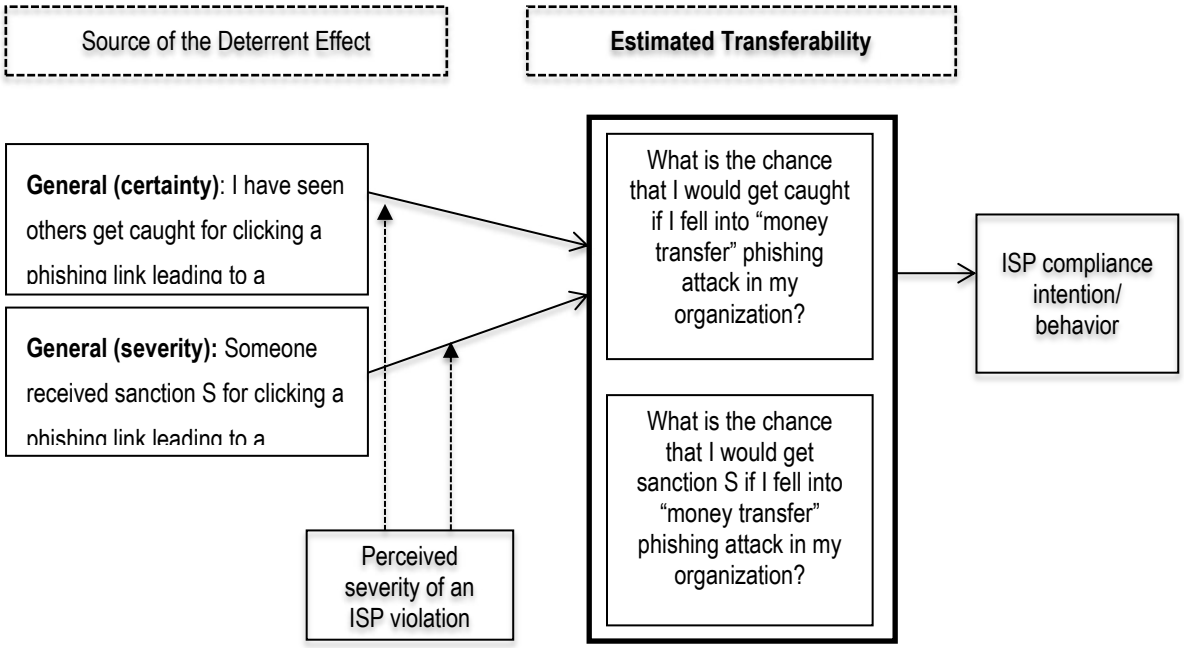

Figure 8. General Deterrent Effects (Certainty and Severity) and their Relationships to Estimated Future Transferability to other ISP Violation with the Perceived Severity of an ISP Violation as a Moderator

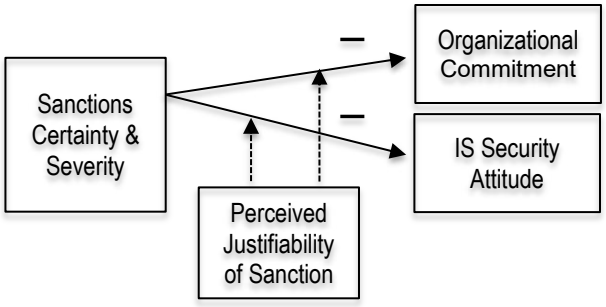

Figure 9. Baseline Model to Examine the Negative Consequences of Unjustified Sanctions, such as Decrease in Organizational Commitment and IS Security Attitude

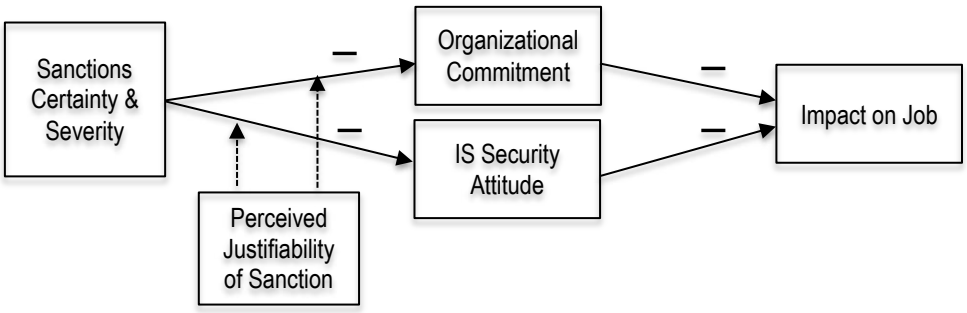

Figure 10. Baseline Model to Examine the Negative Consequences of Unjustified Sanctions, such as Decrease in Organizational Commitment, IS Security Attitude and Impact on Job

Table 1. Core assumptions underlying DT

Assumptions about the ACT	Assumptions about the ACTOR	Assumptions about the PUNISHMENT
1. Criminality of the (punishable) act 2. Social contract underlying the (punishable) act 3. Knowledge of the (punishable) act	4. People are self-interested beings 5. People learn via specific or general deterrence	6. Punishment experience is subjective 7. Punishment for one type of act may deter a different type of act 8. Punishment may backfire

Table 2. The Purpose of Sanctions

Root of Justification	Description	Implications
Backward-looking/ Retribution	• We must punish wrong-doers, even if punishing them diminishes social utility, because justice demands that we punish.	• Regardless of the consequences of the violation, violators must receive the prescribed punishment.
Forward-looking/ Prevention	• We should punish only when doing so would augment social utility. • We must take measures to prevent crimes in the future (again or ever).	• Deterrence implies that crime can be prevented when wrongdoers abstain from crime to avoid being punished. • Incapacitation implies that society is safer because wrongdoers will be made incapable of re-offending.

Table 3. The Assumptions of DT Acknowledged in IS Articles

Assumptions	Recognized (acknowledged)	Empirically applied/studied
The Act		
#1 Criminality of the (punishable) act	**10 articles** (Chen et al., 2018; D'Arcy & Devaraj, 2012; D'Arcy & Herath, 2011; Guo et al., 2011; Lowry et al., 2017; Moody et al., 2018; Theoharidou et al., 2005; Trang & Brendel, 2019; Willison et al., 2018; Zhang et al., 2009).	**3 articles** • 1 article incorporates the assumption in the study design (Guo et al., 2011). • 2 articles invoke the assumption (post-hoc) to explain the non-significant results (Chen et al., 2018; Moody et al., 2018).
#2 Social contract underlying the (punishable) act	**7 articles** (D'Arcy & Herath, 2011; Lee & Lee, 2002; Lee et al., 2004; Siponen et al., 2007; Siponen et al., 2010; Ugrin & Pearson, 2013; Workman & Gathegi, 2007).	n/a
#3 Knowledge of the (punishable) act	**18 articles** (Balozian & Leidner, 2017; Bulgurcu et al., 2010; D'Arcy & Hovav, 2009; D'Arcy et al., 2009; Harrington, 1996; Hovav & D'Arcy, 2007; Knapp et al., 2007; Lee & Lee, 2002; Lee et al., 2004; Lowry et al., 2017; Pahnila et al., 2007; Peace et al., 2003; Siponen et al., 2007; Siponen et al., 2010; Straub, 1990; Straub & Welke, 1998; Theoharidou et al., 2005; Willison et al., 2018).	n/a
The Actor		
#4 People are self-interested beings	n/a	n/a
#5 People learn via specific or general deterrence	**16 articles** (Burns et al., 2017; Choi & Song, 2018; D'Arcy & Devaraj, 2012; D'Arcy & Herath, 2011; Higgins et al., 2005; Kim et al., 2019; Merhi & Ahluwalia, 2019; Moody et al., 2018; Park et al., 2019; Skinner & Fream, 1997; Stoycheff et al., 2019; Straub, 1990; Ugrin & Pearson, 2013; Willison et al., 2018; Wolfe et al., 2008; Xue et al., 2011).	**5 articles** • 1 article incorporates the assumption in the research design (Kim et al., 2019). • 4 articles invoke the assumption to explain non-significant results (Moody et al., 2018; Park et al., 2019; Skinner & Fream, 1997; Ugrin & Pearson, 2013).
The Punishment		
#6 Punishment experience is subjective	**9 articles** (D'Arcy & Herath, 2011; D'Arcy & Hovav, 2009; Hovav & D'Arcy, 2012; Higgins et al., 2005; Hua & Bapna, 2013; Levin et al., 2007; Park et al., 2019; Wolfe et al., 2008).	**3 articles** • 3 articles incorporate the assumption by including the study participants to determine the subjective interpretation of objective sanctions (Higgins et al., 2005; Levin et al., 2007; Wolfe et al., 2008).
#7 Punishment for one act may deter another	n/a	n/a
#8 Punishment may backfire	**12 articles** (Chen & Li, 2014; Chen et al., 2012; Dhillon et al., 2004; Herath & Rao, 2009a; Kim et al., 2019; Knapp et al., 2007; Kuo et al., 2017; Lowry et al., 2015; Rajab & Eydgahi, 2019; Sinha & Mandel, 2008; Stoycheff et al., 2019; Ugrin & Pearson, 2013).	**5 articles** • 5 articles use their empirical findings to show the negative consequences of sanctions (Dhillon et al., 2004; Herath & Rao, 2009a; Rajab & Eydgahi, 2019; Sinha & Mandel, 2008; Stoycheff et al., 2019).
TOTAL	**48 articles** • 31 articles recognized 1 assumption	**15 articles** • 5 articles incorporated 2 assumptions in their study design.

	• 16 articles recognized 2 assumptions • 2 articles recognized 3 assumptions • 1 article recognized 4 assumptions	• 10 articles invoked 3 assumptions to reflect on non-significant empirical findings.

Table 4. Example of Future Research Directions that Clarify the Distinction between the Transferability Assumption and the Two Forms of DT

	General Deterrence	Specific Deterrence
Violations of similar nature (Traditional DT)	To what extent is **observing** someone receive a sanction S for violation X transferable to future IS security violations of the same nature?	To what extent is **personally experiencing** a sanction S for violation X transferable to future IS security violations of the same nature?
Violations of different nature (Transferability)	To what extent is **observing** someone receive a sanction S for violation X transferable to IS security violations of a different nature (e.g., Y and Z)?	To what extent is **personally experiencing** a sanction S for violation X transferable to other IS security violations (e.g. Y and Z)?