Spring 3-31-2020

Deterrence in Cyberspace: An Interdisciplinary Review of the Empirical Literature

David Maimon

Deterrence in Cyberspace: An Interdisciplinary Review of the Empirical Literature

David Maimon

Abstract

The popularity of the deterrence perspective across multiple scientific disciplines has sparked a lively debate regarding its relevance in influencing both offenders and targets in cyberspace. Unfortunately, due to the invisible borders between academic disciplines, most of the published literature on deterrence in cyberspace is confined within unique scientific disciplines. This chapter therefore provides an interdisciplinary review of the issue of deterrence in cyberspace. It begins with a short overview of the deterrence perspective, presenting the ongoing debates concerning the relevance of deterrence pillars in influencing cyber criminals' and cyber attackers' operations in cyberspace. It then reviews the existing scientific evidence assessing various aspects of deterrence in the context of several disciplines: criminology, law, information systems, and political science. This chapter ends with a few policy implications and proposed directions for future interdisciplinary academic research.

Introduction

The considerable literature around the topic of cyber-deterrence continues to grow. Indeed, the popularity of deterrence-based policies in fighting offline crime (Nagin 2013), maintaining diplomatic relationships between countries (Quackenbush 2011), and combating the spread of diseases (Milne et al 2000) has cleared ground for the migration of deterrence-based approaches to cyberspace. In turn, this has sparked a lively debate regarding the relevance of this approach in influencing both cyber attackers' (individuals and countries) malicious and non-malicious online behaviors (Taddeo 2018;Wilner 2019) and targets' online self-protective behaviors (Maimon et al 2017).

Unfortunately, due to the invisible yet rigid boundaries erected between academic disciplines, most of the published literature on deterrence in cyberspace is confined to specific areas and sub-populations which are of limited interest across scientific fields. For example, while criminologists are interested in understanding how sanction threats and punishment influence cybercriminals' behaviors prior to, during the progression of (Maimon et al 2019), and in the culmination of an online criminal event, information systems scholars are more interested in understanding the effectiveness of deterrence-based policies in addressing employees' computer misuse and increasing compliance with their employers' cybersecurity policies (D'Arcy and Herath 2010). Similarly, while law scholars are interested in understanding the necessity of designated substantive cybercrime laws for deterring illegal online activities in the general public and among convicted offenders (Mayer 2015), political scientists tend to focus their debate on the relevance of deterrence-based principles in governing cyber conflicts between nations (Taddeo 2018).

Drawing on the notion that cybercrime research should be of an interdisciplinary nature, generate a comprehensive understanding of relevant concepts in the context of several related fields, and support concrete scientific contributions in each relevant field of study, this chapter intends to provide an interdisciplinary review of the literature around the issue of deterrence in cyberspace. It begins with a short overview of the theoretical premises laid out by deterrence theoreticians, then presents the ongoing debates concerning the relevance of the theory for influencing cyber criminals. The next sections review the existing documented scientific efforts aimed at assessing the validity of different dimensions of deterrence theory, in the disciplinary contexts of criminology, law, information systems, and political science. These efforts a focus on cyber-dependent crimes, i.e. illegal activities that can only be performed using a computer, computer networks, or other forms of information communication technology such as hacking and DDoS attacks (McGuire and Dowling 2013). Finally, the chapter's conclusion proposes a few policy implications and recommends directions for future interdisciplinary academic research.

Deterrence Theory: General Principles

Deterrence theory has its roots in the writings of the 18th-century philosopher Cesare Beccaria (1963 [1764]), who proposed that humans are self-interested and rational decision-makers, driven in their actions by an economical "hedonistic calculus" whereby they seek to maximize pleasure and minimize pain. One key theoretical principle of the theory suggests that individuals are open to "deterrence" inasmuch as raising the costs of a behavior through sanctions would lower their willingness to pursue that course of action. Emphasizing the difference between specific and general deterrence, Beccaria explained that punishments for criminal behaviors aim at both preventing recidivism among convicted criminals (i.e. specific deterrence) and keeping

the general public from engaging in crime (i.e. general deterrence). Ultimately, the theory predicts that while forming expectations regarding the future outcomes of his or her behaviors, an individual's fear of certain, swift, and severe punishment could translate to avoiding criminal behavior altogether (Beccaria 1963 [1764]).

Explaining the delicate balance between severe yet still proportional punishments, Beccaria suggested that punishments should be proportional to the harm inflicted by the criminal act, and that more serious crimes should be followed by more serious punishments. Still, he stressed that it is the certainty of punishment and not its severity that leaves a lasting, deterring impression on the minds of individuals. Accordingly, the certainty of punishment carries a more substantial deterring effect than severe punishment, since the fear of more severe punishment will fail to translate to deterrence if it is accompanied by the hope that one may escape that punishment. As part of his recommendations to criminal justice systems, Beccaria advised authorities to publicize laws (in order to avoid the threat of tyranny), and to make these laws as clear and simple as possible in order to support deterrence efforts. Beccaria's ideas and theoretical principles were conveyed in a classic essay (*On Crimes and Punishment*), which condemned the punitive approaches taken by the Italian criminal justice system during the 18th century when dealing with culprits. Beccaria's essay, along with the work of Bentham (1789) in England, paved the way for a reformation of the early criminal justice systems in Europe and set the stage for the emergence of the criminological field of study.

In parallel to the expansion of deterrence-based policies among global criminal justice agencies and their focus on preventing crime within individuals, the theoretical tenets of the deterrence perspective have also proved useful for guiding sovereign countries' political courses when dealing with rival international players (Jervis 1979). Specifically, Schelling (1960)

suggested that a nation can commit itself to a deterrence strategy that is intended to prevent other nations from opportunistic aggression, by threatening some punishment against potential aggressors and promising rewards for a positive treatment. Explaining the deterring equation further, Snyder argued that "deterrence is a function of the total cost-gain expectations of the party to be deterred, and these may be affected by factors other than the apparent capability and intention of the deterrer to apply punishment or confer rewards" (1961: 9). Credible threats by a deterring party are key in this sense for instilling fear of consequences (Schelling 1966). Deterrence as a coercive national strategy has been discussed in the literature since World War II, yet it started to gain popularity during the 1950s and the Cold War era (Jervis 1979).

Since discussions of cyber-deterrence are relevant both at the individual and group (mainly the state) levels, relevant academic research has been published in multiple academic disciplines, including criminology, law, information systems, and political science. Since their studies are imagined in the context of cyberspace, scholars from all of these fields reflect upon the relevance of deterrence and the applicability of the approach in preventing cyber-dependent crimes. On one hand, several scholars believe that the implementation of deterrence-based strategies (for example, sanctions and sanction threats) in cyberspace is prone to failure, since the inherently anonymous nature of this space complicates the task of attack attribution (Nye 2017) and increases online offenders' ability to escape penalties for their illegitimate online behaviors (Harknett 1996; Harknett et al 2010; Denning and Baugh 2000). This theoretical claim is supported by the notion that potential offenders learn through trial and error that the certainty of being detected and punished for a criminal act is relatively low, so they initiate illegitimate behaviors. Since the certainty of detection and punishment for cyber-dependent crimes is even lower than the certainty of detection of a non-cyber criminal event, due to law enforcement's

lack of preparedness to deal with cyber-dependent crimes (Dupont 2017), the enforcement of sanctions and sanction threats in a computing environment is predicted to play an insignificant role in preventing the occurrence of cyber-attacks (Lupovici 2011).

In contrast, other scholars contend that despite the complexities involved, attribution can still be achieved in cyberspace (Rid and Buchanan 2015; Tor 2017). Still others suggest that it is unnecessary to identify specific individuals in order for deterrence to take effect in cyberspace (Goodman 2010). Accordingly, the introduction of situational deterrence cues in an attacked cyber environment could trigger a predictable avoidance response from an online offender and consequently attenuate the consequences of an online criminal event. For instance, since detection of a system trespassing event results in increased efforts by legitimate users to deny trespassers access to the attacked computer (Stoneburner et al 2002), implementing surveillance measures in a computing environment may lead system trespassers to overestimate the risk of detection on the system, devote increased efforts toward avoiding detection and hiding their presence, and even reduce harmful activity on the system. Therefore, even though deployment of deterrence-based measures in cyber environments will not necessarily prevent the occurrence of online crimes or result in official sanctions, it will increase offenders' efforts to avoid detection and restrict the scope of their illegitimate activity during the progression of an online criminal event.

To test the arguments raised by adherents of these two camps, scholars within these four academic disciplines have tested the validity of deterrence-based arguments in influencing online criminals and their targets. The next section reviews the specific theoretical adjustments scholars have made while using the deterrence perspective to guide academic research, as well as the empirical literature published within each relevant discipline.

Criminological Literature

In his original discussion on the effectiveness of punishment in preventing offenders' subsequent involvement in crime (i.e. recidivism), Beccaria proposed that certain severe and swift punishments would be more effective in deterring criminal behaviors (Beccaria, 1963 [1764]). All in all, findings from extensive criminological research indicate that assigning more severe punishment (i.e. longer prison sentences) carry a modest deterrent effect, and that increasing the certainty of detection and punishment (for example, by deploying more police presence in strategic locations) result in a consistent deterring effect (Nagin 2013). In addition to testing different aspects of classic deterrence in offline environments such as residential neighborhoods (Braga and Weisburd 2012) and schools (Maimon et al 2012), as well as elaborating the difference between general and specific deterrence (Pratt et al 2006), contemporary criminologists have elaborated on different aspects of deterrence, including the impact of punishment avoidance on an individual's decision to initiate a criminal event (Stafford and Warr 1993), the distinction between objective and subjective sanctions (Paternoster 1987), the communication platforms which could be used to convey a coherent deterring message (Geerken and Gove 1974), and the difference between informal and formal sanctions in deterring individuals' involvement in crime (Anderson et al 1977). One additional theoretical elaboration in the context of deterrence was proposed by Gibbs (1975), who differentiated between absolute and restrictive types of deterrence. Gibbs (1975) conceptualized absolute deterrence as an individual's total avoidance of criminal activity due to fear induced by some perceived risk of punishment. Restrictive deterrence, on the other hand, is defined as the (partial) curtailment of a certain type of criminal activity in order to reduce the risk of punishment.

Although deterrence-based research has dominated the criminological discipline within the last five decades, driving numerous investigations of the relationships between key theoretical constructs of deterrence and a wide range of offline crimes (Nagin 1998, 2013), empirical investigations of deterrence-based questions in cyberspace only started to emerge during the late 1990s. Skinner and Fream (1997) investigated the relationships between undergraduate students' perceptions of punishment severity and certainty with their engagement in cybercrime (specifically digital piracy, guessing passwords, manipulating files with no permission, system trespassing, and writing malware). Their findings suggest that students' perceptions of punishment severity were only a significant correlate to system trespassing.

Similarly, Morris and Blackburn (2009) analyzed data collected from a different sample of undergraduate students. These scholars reported that a measure tapping students' assessment of the chances of getting caught and their perceptions of severe punishment was significantly associated with password guessing, attempted hacking, and file manipulation. However, the theoretical framework that guided these two studies was that of social learning theory (Akers 2017). Moreover, although these studies offer preliminary investigations of the relationship between an individual's perception of punishment severity and certainty and his or her involvement in various online crimes, they still leave something to be desired due to the questionable operationalization of key deterrence constructs (especially in Morris and Blackburn 2009) and their low alpha scores, the student-based sample they rely upon, and the cross sectional nature of the data.

More recently, Holt and colleagues (2017) reported a significant association between students' perception of law-enforcement's likelihood to quickly recognize a cybercrime event and their willingness to engage in an ideologically motivated cyber-attack against a foreign

country. However, this research suffers from similar problems to those observed in Skinner and Fream (1997) and Morris and Blackburn (2009). In fact, Holt and associates' reliance on students' responses to vignettes for constructing their dependent variables is problematic in the context of deterrence, since the subjects' lack of realistic understanding of the true costs (and benefits) of hacking casts a shadow on the validity of the constructs they create. Taking a somewhat different approach toward investigating the relationship between law enforcement reports of detection of online crime events and cybercrime, Guitton (2012) collected data on the number of attacks reported against businesses in France, Germany, and the UK between the years 2003 and 2010, then correlated the data with several proxies for law enforcements' successful operations in cyberspace. Findings from his analyses suggest that the rate of newspaper articles reporting cybercrime incidents with a lack of attribution is positively related to the number of cyber-attacks reported against businesses in each of the observed countries. Given the serious methodological difficulties embedded in Guitton's approach to data collection, any conclusion drawn regarding the effectiveness of attribution should be taken cautiously.

Although early criminological research mainly employed survey designs and student-based samples to explore the relationships between deterrence-based constructs and cybercrime, several studies have investigated whether different aspects of deterrence influence the progression of cyber-dependent crimes using experimental research designs (Maimon et al 2014, Wilson et al 2015, Testa et al 2017, Maimon et al 2019). These studies adopted Gibbs' (1975) conceptualization of restrictive deterrence to guide their efforts in assessing the effectiveness of deterrence-based interventions in shaping the progression of system trespassing events. Maimon and colleagues (2014), for example, tested the effect of a warning banner in an attacked computer system on the progression, frequency, and duration of system trespassing events.

Deploying a large set of target computers built for the sole purpose of being attacked (i.e., honeypots) on the Internet infrastructure of a large USA university, these scholars revealed that although a warning banner did not lead to the immediate termination of trespassing incidents or reduce their frequency, it did result in a shorter average duration of the system trespassing incidents. Interestingly, the effect of a warning message on the duration of repeated trespassing incidents was attenuated in computers with a large bandwidth capacity. Stockman et al (2015) offered further support for these findings.

Testa and colleagues (2017) explored the effect of a warning banner in mitigating hackers' levels of activity (i.e., roaming the attacked system and manipulating file permissions) in an attacked computer system, while considering the level of administrative privileges imposed by the system trespasser on the attacked computer. Analyzing data collected by Maimon and colleagues (2014) in their second experiment, Testa and associates (2017) reported that the presence of a warning banner on an attacked computer system had no statistically significant effect on the probability of either navigation or file permission change commands being entered on the system. However, when testing the effect of the warning banner on computers attacked by system trespassers with non-administrative privileges, the authors reported that a warning banner substantially reduced the use of both navigation and change file permission commands, compared to the no-warning computers. More recently, Maimon et al (2019) analyzed data collected in a randomized trial which was deployed in China, reporting that intruders are less likely to use "clean tracks" commands in the presence of detection by a legitimate user of an attacked computing environment, in the absence of subsequent presentation of sanction threats.

In addition to investigating the effectiveness of sanction threats in deterring the progression of hacking incidents, several scholars have investigated the effect of surveillance and

detection signs in restricting the scope of hackers' illegitimate behaviors while taking over a system. Wilson and associates (2015), for example, assessed the effect of a surveillance banner on the probability of commands being entered in the attacked computer system. They found that the presence of a surveillance banner in the attacked computer systems reduced the probability of commands being typed in the system during longer initial system trespassing incidents. Further, they reported that the probability of commands being typed during subsequent system trespassing incidents (on the same target computer) was conditioned by the presence of a surveillance banner and by whether commands had been entered during previous trespassing incidents. Using the same data, Maimon and colleagues (2019) investigated whether the level of ambiguity regarding the presence of surveillance in an attacked computer system influences system trespassers' likelihood to clean their tracks during the progression of an event. Their findings indicate that the presence of unambiguous signs of surveillance (i.e. the presence of both a surveillance banner and program in the attacked system) increases the probability of clean tracks commands being entered on the system.

Despite the growing use of honeypots for understanding system trespassers' behaviours during the progression of criminal event among criminologists (Maimon et al 2014; Wilson et al 2015) and computer scientists (Farinholt et al 2017; Rezaeirad et al 2018), these tools present some methodological challenges to scholars (Holt 2017). For starters, while these simulated environments are indistinguishable from standard legitimate devices for less sophisticated hackers, fingerprinting techniques can be used by hackers to distinguish between regular online environments and honeypots (Mohammadzadeh, Mansoori, and Welch, 2013). In addition, honeypots are able to measure explicit actions but are unable to measure the fundamental attitudes, beliefs, and capabilities of intruders who interact with the honeypot. Finally, honeypots

are also unable to detect communications such as warnings and recommendations between hackers that may alter behaviour within a honeypot (Holt 2017). Still, the usefulness of honeypots in understanding system trespassers responses to various computer configurations during the progression a criminal event is unique, and these findings should guide the design of more secure computing environments.

Law Literature

Substantive criminal laws set behavioral standards for individuals in society, detail legal rules that forbid specific types of behaviors, and elaborate potential legal sanctions imposed for deviating from these laws. Since cybercrime has become a serious threat to individuals, organizations, and governments all around the world, many countries have realized the necessity of establishing an arsenal of well-defined cybercrime laws which can guide law enforcement agencies' efforts to pursue online offenders (Brenner 2001), and thus have enacted laws that prohibit specific types of behaviors with computers and computer networks. For example, in the USA, the Computer Fraud and Abuse Act 2008 prohibits accessing a computer without authorization or using a computer to defraud and extort. Similarly, in the UK, the Computer Misuse Act 1990 makes it illegal to gain improper access to a computer or to commit theft and extortion using computers. China, on the other hand, has amended the relevant provision of its criminal code twice (Amendment VII to the Criminal Law of the People's Republic of China, 2009) to prevent illegal online activities. Importantly, in addition to cybercrime-specific laws, many countries also rely on laws designed to prevent terrestrial crimes when targeting certain types of cybercriminals.

The underlying premise behind the enactment and enforcement of official cybercrime laws draws on the assumption that rational humans will be deterred from engagement in

illegitimate online activities once threatened by harsh, immediate, and certain punishments for initiating these behaviors. However, although extensive research has explored the effectiveness of familiarity with laws and the administration of criminal justice procedures across different junctions of the criminal justice system, in deterring individuals' onset of criminal career and recidivism (Paternoster 2010), we know little about the effectiveness of cybercrime laws in preventing cybercrime incidents. Still, recent evidence on the effectiveness of USA cybercrime laws in preventing online crimes is starting to emerge.

Mayer (2015), for example analyzed data from hundreds of civil and criminal pleadings that were processed in the USA between the years 2005-2012, and proposed that the *Computer Fraud and Abuse Act* (CFAA)[1] cannot deter cyber criminals. Accordingly, since the potential benefits from initiating a cybercrime incident during those years outpaced the potential punishment enforced in both criminal and civil cases, the deterrence benefits of this law are negligible. Similarly, in a series of papers, Kigerl (2009, 2016, 2015, 2018) explored the potential impacts of the *Controlling the Assault of Non-Solicited Pornography and Marketing Act* (CAN SPAM Act)[2] of 2013 on different aspects of email spamming. Analyzing data collected using multiple "honey-net" email addresses posted online for spammers to find and send spam emails to, Kigerl reported mixed findings regarding the deterring effect of this act. For example, while the CAN SPAM act had no effect on the amount of spam sent to targets (Kigerl 2009, 2016), or on the probability that spammers would embed their physical address in the spam email, the enactment of this act is positively associated with adding a verbal description of the email in the email's subject line.

[1] The CFAA aims to prevent unauthorized access to computers and password trafficking.
[2] CAN SPAM Act aims to regulate the way unsolicited commercial emails are sent to email users, and to regulate the content that the email messages deliver.

Integrating the honey-net data with data collected from news articles published on the topic of CAN SPAM act, Kigerl (2016) reported that the number of ongoing CAN SPAM trials reported in popular news outlets is associated with a reduction in the amount of spam email sent, particularly in the USA (Kigerl 2018). In contrast, the volume of news articles reporting spammers' detention seem to be positively associated with the volume of spam (Kigerl 2016), as well as with violation of email header forgery laws (Kigerl 2015). Finally, Hui and colleagues (2017) investigated the potential deterring effect of the 2001 *Convention on Cybercrime* (COC)[3] legislation, in terms of reducing the volume of DDoS attacks against enforcing countries. Analyzing data on DDoS attacks reported in 106 countries during 177 days between 2004 and 2008, these authors reported that enforcing the COC decreased DDOS attacks by at least 11.8%. Moreover, Hui and associates (2017) observed that enforcement of the COC resulted in an increase of the number of DDoS attacks against non-enforcing countries.

Although key for generating an initial understanding regarding the effect of cybercrime laws in deterring online criminal activities, the problematic nature of the samples and data employed throughout the studies reported in this section should be considered carefully. First, scholars' reliance on secondary data and unfamiliarity with the full extent of the methodology behind the original data collection may raise questions regarding the validity of some of the constructs composed by the scholars. Moreover, drawing on news websites and popular media to construct key independent measures may introduce some selection bias, since only some cases end up in the media. Finally, failure to control for internal processes that occur at the organized-

[3] The Convention on Cybercrime is an international treaty on crimes committed over the internet which seeks to improve international cooperation between nations in cybercrime investigations and harmonize national cybercrime laws.

crime group level, and that may influence the volume of online crime (Krebs 2014), calls into question the findings reported in these papers.

Information Systems Literature

Consistent with the criminological literature which focuses on understanding different aspects of punishment in preventing online crime, empirical attention has been devoted within the Information Systems field to exploring different aspects of sanctions in preventing cyber-dependent crimes. However, while criminologists focus on online offenders, Information System scholars mostly aim to understand computer misuse by employees in organizations, as well as employees' compliance with organizational security policies (Cram et al 2017). In general, findings regarding the effectiveness of sanctions in reducing employees' computer misuse and violation of information security policies are mixed (D'Arcy and Herath 2010). For example, although some research reports that punishment severity decreases intentions to violate information security policies, technology misuse, and computer abuse (D'Arcy et al 2009; Devarj 2012; Cheng et al 2013), other studies find this effect in the USA only (Hovav and Darcy 2012) while still others do not observe this relationship at all (Hu et al 2011). Moreover, the effect of sanctions' certainty in reducing intentions to misuse information security was significant for specific populations only (Darcy et al 2009; Hovav and Darcy 2012). Finally, several studies find an insignificant effect of sanction celerity on individuals' violation of information security policies (Hu et al 2011). Still, Barlow and colleagues (2013) observed that clearly communicating sanctions is key for reducing intentions to violate information security policies among employees.

Since several scholars believe that non-compliance and violations of security policies are behaviors distinct from compliance behaviors (Guo 2013), extensive research has also assessed

the role of deterrence in encouraging employees' compliance with organizational cybersecurity policies. In fact, Sommestad and colleagues' (2014) systematic review of the key variables that influence information security compliance behaviors suggested that deterrence-based sanctions are stronger predictors of policy compliance than non-compliance. Herath and Rao (2009a), for example, reported that sanction certainty increases employees' intention to comply with information security policy. Operationalizing subjects' assessments of detection probability as a proxy for sanction certainty, Li and associates (2010) further confirmed this finding. However, both studies failed to observe a significant relationship between perception of punishment severity and compliance with organizational security policies. Chen and associates (2012) reported that in addition to the effectiveness of punishment certainty, employees' high certainty for rewards increased their intentions to comply with information security policy.

In addition to exploring the effects of deterrence-based strategies on employees' computer abuse and compliance/non-compliance with security policies, extensive IS research has investigated ways in which rewards and punishments could influence employees' decisions to engage in self-protective behaviors (Herath and Rao 2009b, Johnston and Warkentin 2010, Siponen et al 2010). This line of research draws on Protection-Motivation Theory (PMT) (Rogers 1975, 1983), which suggests that individuals are more likely to protect themselves from potential risks after receiving fear-arousing recommendations. Specifically, two processes must occur for a person to engage in an adaptive protective response. First, in the threat-appraisal process, the threat and generated fear that inspire protection motivation must be weighted more heavily than the maladaptive rewards earned by not engaging in protection motivation. Second, in the coping-appraisal process, a person's response efficacy and self-efficacy must outweigh the response costs for engaging in the protection motivation (Rogers 1975).

Consistent with samples used to investigate deterrence-based premises in the Information Systems field, most of the empirical research employing PMT draws on data collected from samples of organizational employees. Herath and Rao (2009b) reported that employees make inaccurate predictions about the probability of experiencing a security breach in their organizations, which in turn resulted in non-compliance with organizational security policies. In contrast, employees' accurate assessments of their organizational vulnerability to information security threats was found to have a significant effect on their intentions to comply with security policies (Siponen et al 2010). Moreover, Workman and colleagues (2008) reported that employees' subjective assessments of risk severity as a result of a breach of their confidential information, as well as their perceived vulnerability to cyber-dependent crime, were negatively associated with failure to apply security solutions. Finally, focusing on the relationships between fear appeals and the enactment of computer security behaviors, Johnson and Warkentin (2010) reported that while there is an overall positive effect of fear appeal on the use of computer security behaviors, this effect varies in magnitude across users and based on individuals' personality traits (for example level of self-efficacy), cognitive processes (i.e. threat severity), and social influence (see also Boss et al 2015 and Siponen et al 2010). Still, no prior research has explored whether employees' compliance with organizational security policies reduces the organization's risk of cyber-dependent crime victimization.

Similar to the issues embedded in survey-based criminological research of online offenders, the bulk of the Information Systems scholarship which focuses on deterrence suffers from various methodological issues. Specifically, the focus on employees' intentions instead of actual illegitimate activities with their organizational networks (Li et al 2010), questionable operationalizations of key deterrence constructs (Siponen et al 2010), and the cross-sectional

nature of the data collected and reported in most of these studies raise questions regarding the observed empirical patterns.

Political Science Literature

In contrast to criminologists', information systems scientists', and law scholars' interest in exploring how different aspects of deterrence determine individual involvement in online crime, political scientists' discussions on cyber-deterrence are focused on countries' efforts to prevent and dissuade rival nations' attempts to launch cyber-attacks. Specifically, Libicki (2009) suggested that the goal of cyber-deterrence is to attenuate the risk of cyber-attacks to an acceptable level at an acceptable cost, when a defending state aims to mitigate potential offensive actions by threatening a potential retaliation. Several scholars identify the means thorough which nations' deterring postures in cyberspace could be achieved. Iasiello (2014), for example, differentiated between *deterrence by punishment* and *deterrence by denial* (see also Nye (2016) and Lupovici (2011)). Specifically, while deterrence by punishment is focused on conveying to potential attackers that significant sanctions will be imposed in retaliation to any cyber-attack, deterrence by denial aims to convey to potential attackers that their aggressive efforts in cyberspace will be futile. Importantly, Iasiello (2014) argued that the key factors required for supporting both means of deterrence are (1) effective communication of the deterring messages, (2) the ability to properly signal intentions to receivers, (3) the ability to successfully attribute attacks to an aggressor, and (4) proportional retaliation for different cyber-attacks. Nye (2017) identified two additional means of deterrence: *entanglement* and *norms*. Entanglement refers to the presence of interdependencies which make the consequence of an attack serious to both the attacker and the target. Similarly, normative considerations refer to the potential reputational costs that may follow a cyber-attack and which may damage an actor's soft

power beyond the value gained from an attack. Finally, Tor (2017) discussed the relevance of *cumulative deterrence* against cyber-attacks. Drawing on the rationale advanced by Gibbs (1975) in his discussion of restrictive deterrence, cumulative deterrence refers to repeated attacks on a rival in response to specific behaviors, over a long period of time, and in some cases disproportionally to the attacker's aggressive behaviors (Tor 2017).

Given the considerable theoretical attention provided in the political science discipline to the different means of cyber-deterrence, one may expect a similar level of scientific exploration around the effectiveness of various deterrence approaches in preventing and mitigating nations' aggression in cyberspace. However, a recent systematic review by Gorwa and Smeets (2019) suggested that such empirical works are still missing in this field. Indeed, only two studies (Kostyuk and Zhukov 2019; Valeriano and Maness 2014) examined the dynamic of cyber conflict between rival nations while employing quantitative research designs methods. Specifically, Valeriano and Maness (2014) analyzed data from 110 cyber incidents and 45 cyber disputes and found that when cyber operations and incidents occur, they tend to carry a minimal impact and low severity due to the dynamic of cyber restraint. In contrast, Kostyuk and Zhukov's (2019) analyses of cyber-attack data collected during the conflicts in Ukraine (between 2013-2016) and Syria (between 2011-2016) revealed that cyber-attacks do not facilitate an effective vehicle of coercion during war. Unfortunately, neither of these papers examine the theoretical aspects of cyber-deterrence and their effectiveness in preventing and dissuading cyber-attacks.

Policy Implications and Directions for Future Research

The Comprehensive National Cybersecurity Initiative (CNCI) has evolved to become one of the central elements of U.S. national cyber security strategy (www.whitehouse.gov). One key activity in the CNCI highlights the development of deterrence-based strategies designed to

prevent and mitigate the consequences of cyber-attacks against U.S. organizations and individuals. However, despite the emphasis placed on cyber-deterrence, this review reveals mixed evidence regarding the effectiveness of different aspects of deterrence-based strategies in preventing the occurrence of malicious cyber activities. Specifically, studies published in both criminological and information systems journals suggest that the effect of sanction severity in preventing online crime is inconsistent (Skinner and Fream 1997; Morris and Blackburn 2009; D'Arcy et al 2009; Hovav and D'Arcy 2012), and that the effect of punishment certainty is only significant among specific populations of online offenders (Morris and Blackburn 2009; Darcy et al 2009; Hovav and Darcy 2012). In contrast, the detection and attribution of online crime (Guitton 2012; Maimon et al 2019), along with clear communication of sanctions (Barlow et al 2013), are consistently found to be negatively associated with online criminals' willingness to launch cyberattacks and adopt avoidance strategies. Moreover, the effectiveness of various deterrence-based methods in restricting the scope and disrupting the progression of online criminal events has been reported in several criminological studies (Maimon et al 2014; Wilson et al 2015; Testa et al 2017).

Finally, while previous information systems and law research tends to test the effect of general deterrence (although not explicitly) on online crime (Kigerl 2016; Mayer 2015; Hui et al 2017), no prior research has tested the effect of punishment on online offenders' recidivism. Therefore, in addition to ongoing policy efforts which aim to prevent online crime by adopting deterrence-based policies, practitioners and cybersecurity experts should consider adopting deterrence-based approaches for mitigating the consequence of online criminal events (Maimon and Louderback 2019; Willison, Lowry and Paternoster 2018). This approach should also guide cybersecurity experts' efforts to develop new technical tools which may support the mitigation

and discovery of cybercrime incidents. Indeed, these experts have devoted considerable attention in the last twenty years to developing tools that are designed to detect computer and network vulnerabilities, and to prevent cybercrimes from developing (Waldrop 2016). Although such tools are designed to identify vulnerabilities and prevent their exploitation by malicious actors, none of them allow rapid detection of these incidents or effective mitigation of the consequences of an attack. Moreover, the effectiveness of these tools in preventing online crime is questionable. Therefore, configuring new tools while drawing on the restrictive deterrence approach may prove useful in reducing the scope of online offenders during the progression of cybercrime events (Gibbs 1975).

Future research should further explore whether cyber-deterrence prevents and disrupts the progression of cyber-crimes. Such research should explore the influence of both absolute and restrictive cyber-deterrence. As elsewhere, one of the major hurdles in this area could be the absence of universally accepted measurement metrics, which would provide guardians and scholars with practical techniques for assessing the effectiveness of deterrence-based efforts, security policies, and tools in preventing cybercrime (Torres et al 2006). Indeed, the most common approach to the implementation of preventive practices in online environments draws on guardians' personal experience in the field, as well as their personal world views when making security-related decisions that may influence offenders and targets (Siponen and Willison 2009). Such an approach does not require rigorous empirical evaluations of security tools and policies to support the decision-making by these professionals. In fact, Blakely (2002) suggested that this approach has failed to prevent individuals and organizations from becoming the targets and victims of cybercrime. Therefore, Blakely proposed the adoption of an approach

that monetizes guardianship efforts and quantifies the effectiveness of security tools and policies in achieving their stated goals.

Scholars within each of the scientific disciplines reviewed in this paper should consider conducting empirical research that will push the envelope in the context of all four disciplines simultaneously. Criminologists and information systems scholars should seek to collect data on online crime directly from the field and create better operationalization of deterrence-based concepts. Given the cynicism developed in the criminological field about the collection and analysis of data, datasets that are used by cyber criminologists in their publications should be made publicly available. Future criminological research should further explore how different configurations of online environments shape both online offenders' and targets' involvement in cyber-dependent crimes (Lessig 2009). Encouraged by findings reported in the criminological literature indicating that environmental design could reduce the volume of robbery (Jeffrey et al 1987), vandalism (Sloan-Howitt and Kelling 1990), and shoplifting (Farrington and Burrows 1993) incidents, guardians' familiarity with computer and online configurations that result in lower rates of and less damage from cyber-dependent crimes could guide the design of safer online environments.

Law scholars should attempt to produce empirical assessments regarding the effectiveness of computer crime laws in deterring crime, both in the USA and in other places around the globe. Particular attention should be given to investigating the effectiveness of official punishment in reducing recidivism among convicted online criminals. Similarly, political scientists should also seek to produce more empirical research around the effectiveness of cyberwarfare in preventing and mitigating nations' aggression in cyberspace (Gorwa and Smeets 2019). Furthermore, future law research should continue to assess the effect of international

collaborations on cybercrime laws and their enforcement in reducing different types of online crime around the world.

Finally, future research should seek to evaluate the most effective ways to successfully implement deterrence-based policies and law enforcement operations in online environments, as well as to assess the effectiveness of these approaches in preventing and mitigating the consequences of cybercrime. Such evaluations should include the development of cybercrime metrics that are clear, objective, repeatable, and simple (Atzeni and Lioy 2006).

Conclusions

Governmental agencies and private corporations around the globe employ a wide range of cyber laws, technical tools, and security policies in efforts to reduce their probability of becoming victims of cybercrime. Many of the laws, security tools, and organizational procedures utilize deterrence-based strategies, which aim to prevent the occurrence of offline crimes by threatening potential offenders with sanctions. Unfortunately, despite the prevalence of cyber-deterrence policies and strategies, the effectiveness of deterrence-based strategies in preventing and mitigating the occurrence of online crime is still relatively unknown within the criminological, law, information systems, and political science fields. Therefore, interested scholars within these academic disciplines should seek to produce rigorous evidence regarding deterrence-based policies, sanctions, and threats in preventing and mitigating cybercrime events. Specifically, while the available evidence regarding the effectiveness of cyber-deterrence in preventing cybercrime tends to draw mainly on survey-based research, efforts should be made to conduct scientific research that investigates the effectiveness of swift, severe, and certain sanctions for online crime in the wild, through the implementation of experimental research designs. This should be done while paying close attention to the immense disconnect that exists between the

various academic disciplines around the topic of cyber-deterrence. Active efforts should seek to bridge this disconnect, in order to allow a more comprehensive and thorough understanding of different aspects of deterrence in cyberspace.

References

Akers, R. (2017). *Social learning and social structure: A general theory of crime and deviance*. Routledge.

Anderson, Linda S., Theodore G. Chiricos, and Gordon P. Waldo. 1977. "Formal and informal sanctions: A comparison of deterrent effects." *Social problems* 25, no. 1: 103-114.

Barlow JB, Warkentin M, Ormond D, Dennis AR. 2013. Don't make excuses! Discouraging neutralization to reduce IT policy violation. *Computers and Security* 39:145-59.

Beccaria, Cessare. (1963). On crimes and punishments (H. Paolucci, Trans.). *Indianapolis, IN: Bobbs-Merrill.(Original work published 1764)*.

Bentham, J. 1789. The Principles Of Morals and Legislation. Prometheus Books.

Braga, Anthony A., and David L. Weisburd. "The effects of focused deterrence strategies on crime: A systematic review and meta-analysis of the empirical evidence." *Journal of Research in Crime and Delinquency* 49, no. 3 (2012): 323-358.

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

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

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* 39:447-59.

Cram, W. A., Proudfoot, J. G., and D'Arcy, J. (2017). Organizational information security policies: a review and research framework. *European Journal of Information Systems*, *26*(6), 605-641.

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 Res.* 20: 79-98.

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 J. of Information Systems* 20: 643-58.

Denning, D., Baugh, W.: Hiding crimes in cyberspace. In: Thomas, D., Loader, D. (eds.) Cybercrime: Law Enforcement, Security and Surveillance in the Information Age, pp. 105–132. Routledge, London (2000)

Dupont, B. 2017. "Bots, Cops, and Corporations: On the Limits of Enforcement and the Promise of Polycentric Regulation as a Way to Control Large-Scale Cybercrime." *Crime, Law, and Social Change 67*: 97-116.

Farrington DP, Burrows JN. 1993. Did shoplifting really decrease?. *The British Journal of Criminology* 33:57-69.

Farinholt, B., Rezaeirad, M., Pearce, P., Dharmdasani, H., Yin, H., Le Blond, S., McCoy, D. and Levchenko, K. 2017. To catch a ratter: Monitoring the behavior of amateur darkcomet rat operators in the wild. In *2017 IEEE Symposium on Security and Privacy (SP)* (pp. 770-787).

Geerken, Michael R., and Walter R. Gove. 1974. "Deterrence: Some theoretical considerations." *Law and Soc'y Rev.* 9: 497.

Gibbs, J.1975. Crime, Punishment, and Deterrence. Elsevier Scientific Publishing Company, New York

Goodman, W. 2010. Cyber-deterrence: tougher in theory than in practice? Strategic Studies Quarterly Fall, pp. 102–135

Gorwa, R., and Smeets, M. (2019). Cyber Conflict in Political Science: A Review of Methods and Literature.

Guitton, C. (2012). Criminals and Cyber Attacks: The Missing Link Between Attribution and Deterrence. *International Journal of Cyber Criminology*, *6*(2).

Guo, K. H. (2013). Security-related behavior in using information systems in the workplace: A review and synthesis. *Computers and Security*, *32*, 242-251.

Harknett, R.1996. Information warfare and deterrence. Parameters 26, 93–107

Harknett, R., Callaghan, J., Kauffman, R.: Leaving deterrence behind: war-fighting and national cybersecurity. J. Homel. Secur. Emerg. Manag. 7(1), 1–24 (2010)

Herath, T., and Rao, H. R. (2009a). Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. *Decision Support Systems*, *47*(2), 154-165.

Herath T, Rao HR. 2009b. Protection motivation and deterrence: a framework for security policy compliance in organisations. *European J. of Information Systems* 18: 106-25.

Holt, T. J. 2017. On the value of honeypots to produce policy recommendations. *Criminology and Public Policy*, *16*(3), 739-747.

Holt, T. J., M. Kilger, L. Chiang, and C. Yang. 2017. "Exploring the Correlates of Individual Willingness to Engage in Ideologically Motivated Cyberattacks." *Deviant Behavior 38*: 356-373.

Hovav A, D'Arcy J. 2012. Applying an extended model of deterrence across cultures: An investigation of information systems misuse in the US and South Korea. *Information and Management* 49:99-110.

Hu Q, Xu Z, Dinev T. and Ling H. 2011. Does deterrence work in reducing information security policy abuse by employees?. *Communications of the ACM* 54:54-60.

Hui, K. L., Kim, S. H., and Wang, Q. H. (2017). Cybercrime deterrence and international legislation: Evidence from distributed denial of service attacks. *Mis Quarterly*, *41*(2), 497.

Iasiello, E. (2014). Is cyber-deterrence an illusory course of action?. *Journal of Strategic Security*, *7*(1), 54-67.

Jeffrey, CR., Hunter, RD. and Griswold J. 1987. Crime prevention and computer analysis of convenience store robberies in Tallahassee. *Florida Police Journal* 34: 65–69.

Jervis, Robert. 1979. "Deterrence theory revisited." *World Politics*31, no. 2: 289-324.

Johnston, AC, M. Warkentin. 2010. Fear Appeals and Information Security Behaviors: An Empirical Study. *MIS Quarterly* 34: 549-566.

Kigerl, A. C. (2009). CAN SPAM Act: An empirical analysis. *International Journal of Cyber Criminology*, *3*(2), 566.

Kigerl, A. C. (2015). Evaluation of the CAN SPAM ACT: Testing deterrence and other influences of e-mail spammer legal compliance over time. *Social Science Computer Review*, *33*(4), 440-458

Kigerl, A. C. (2016). Deterring spammers: impact assessment of the CAN SPAM act on email spam rates. *Criminal Justice Policy Review*, *27*(8), 791-811.

Kigerl, A. C. (2018). Email spam origins: does the CAN SPAM act shift spam beyond United States jurisdiction?. *Trends in Organized Crime*, *21*(1), 62-78

Kostyuk, N., and Zhukov, Y. M. (2019). Invisible digital front: Can cyber attacks shape battlefield events?. *Journal of Conflict Resolution*, *63*(2), 317-347.

Krebs, B. (2014). *Spam nation: the inside story of organized cybercrime-from global epidemic to your front door*. Sourcebooks, Inc..

Lessig L. 2009. *Code 2.0.* Seattle, WA: Amazon CreateSpace Publishing..

Li, H., Zhang, J., and Sarathy, R. (2010). Understanding compliance with internet use policy from the perspective of rational choice theory. *Decision Support Systems*, *48*(4), 635-645.

Libicki, M. C. (2009). *Cyberdeterrence and cyberwar*. Rand Corporation.

Lupovici, A. (2011). Cyber warfare and deterrence: trends and challenges in research. *Military and Strategic Affairs*, *3*(3), 49-62.

Maimon, David, Olena Antonaccio, and Michael T. French. 2012 "Severe sanctions, easy choice? Investigating the role of school sanctions in preventing adolescent violent offending." *Criminology* 50, 2 : 495-524.

Maimon, D., M. Alper, B. Sobesto, and M. Culkier. 2014. “Restrictive Deterrent Effects of a Warning Banner in an Attacked Computer System.” *Criminology 52*: 33-59.

Maimon, D., Becker, M., Patil, S., and Katz, J. 2017. Self-Protective Behaviors Over Public WiFi Networks. In *The {LASER} Workshop: Learning from Authoritative Security Experiment Results ({LASER} 2017)* (pp. 69-76).

Maimon, D., Testa, A., Sobesto, B., Cukier, M. and Ren, W., 2019, July. Predictably Deterrable? The Case of System Trespassers. In *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage* (pp. 317-330). Springer, Cham.

Milne, S., Paschal Sheeran, and Sheina Orbell. 2000. "Prediction and intervention in health-related behavior: A meta-analytic review of protection motivation theory." *Journal of Applied Social Psychology* 30, no. 1: 106-143.

Mohammadzadeh, H., Mansoori, M., and Welch, I. (2013, January). Evaluation of fingerprinting techniques and a windows-based dynamic honeypot. In *Proceedings of the Eleventh Australasian Information Security Conference-Volume 138* (pp. 59-66). Australian Computer Society, Inc..

Morris, R. G., and Blackburn, A. G. (2009). Cracking the code: An empirical exploration of social learning theory and computer crime. *Journal of Crime and Justice*, *32*(1), 1-34.

Nagin, Daniel S. 1998. "Criminal deterrence research at the outset of the twenty-first century." *Crime and justice* 23, 1-42.

Nagin, Daniel S. 2013. "Deterrence: A review of the evidence by a criminologist for economists." *Annual Review of Economy* 5, no. 1: 83-105.

Nye Jr, J. S. 2017. Deterrence and dissuasion in cyberspace. *International Security*, *41*(3), 44-71.

Paternoster, Raymond. 1987. "The deterrent effect of the perceived certainty and severity of punishment: A review of the evidence and issues." *Justice Quarterly* 4, no. 2: 173-217.

Paternoster, R. 2010. How much do we really know about criminal deterrence. *Journal of Criminal Law and Criminology*, *100*, 765.

Pratt, T. C., Cullen, F. T., Blevins, K. R., Daigle, L. E., and Madensen, T. D. 2006. The empirical status of deterrence theory: A meta-analysis. *Taking stock: The status of criminological theory*, *15*, 367-396.

Quackenbush, Stephen L.2011. "Deterrence theory: where do we stand?." *Review of International Studies* 37, no. 2: 741-762.

Rid, Thomas, and Ben Buchanan. 2015. "Attributing cyber attacks." *Journal of Strategic Studies* 38, no. 1-2: 4-37.

Rezaeirad, M., Farinholt, B., Dharmdasani, H., Pearce, P., Levchenko, K. and McCoy, D., 2018. Schrödinger's {RAT}: Profiling the Stakeholders in the Remote Access Trojan Ecosystem. In *27th {USENIX} Security Symposium ({USENIX} Security 18)* (pp. 1043-1060).

Rogers RW. 1975. A protection motivation theory of fear appeals and attitude change. *J. of Personality* 91:93-114.

Rogers, R. W. (1983). Cognitive and psychological processes in fear appeals and attitude change: A revised theory of protection motivation. *Social psychophysiology: A sourcebook*, 153-176.

Schelling, Thomas C. *The strategy of conflict*. Harvard university press, 1980.

Schelling, Thomas C. "Arms and Influence Yale University Press." *New Haven* (1966).

Siponen M, Pahnila S, Mahmood MA. 2010. Compliance with information security policies: An empirical investigation. *Computer*. 43:64-71.

Skinner, W. F., and A. M. Fream. 1997. "A Social Learning Theory Analysis of Computer Crime among College Students." *Journal of Research in Crime and Delinquency 34*: 495-518.

Sloan-Howitt M, Kelling GL. 1990. Subway graffiti in New York City: Gettin'up vs. meanin'it and cleanin'it. *Security J.* 1:131-6

Snyder, Glenn Herald. 1961. *Deterrence and defense*. Princeton University Press.

Sommestad, T., Hallberg, J., Lundholm, K., and Bengtsson, J. (2014). Variables influencing information security policy compliance: a systematic review of quantitative studies. *Information Management and Computer Security*, *22*(1), 42-75.

Stafford, Mark C., and Mark Warr. "A reconceptualization of general and specific deterrence." *Journal of research in crime and delinquency* 30, no. 2 (1993): 123-135.

Stockman, M., Heile, R., and Rein, A. 2015. An open-source honeynet system to study system banner message effects on hackers. In *Proceedings of the 4th Annual ACM Conference on Research in Information Technology* (pp. 19-22).

Stoneburner, G, Goguen, A, Feringa, A .2002. Risk Management Guide for Information Technology Systems. *NIST Special Publication* 800:30.

Taddeo, M. 2018. The limits of deterrence theory in cyberspace. *Philosophy and Technology*, *31*(3), 339-355.

Testa, A., D. Maimon, B. Sobesto, and M. Cukier. 2017. "Illegal Roaming and File Manipulation on Target Computers: Assessing the Effect of Sanction Threats on System Trespassers' Online Behaviors." *Criminology and Public Policy 16*: 687-724.

Tor, U. (2017). 'Cumulative Deterrence as a New Paradigm for Cyber-deterrence. *Journal of Strategic Studies*, *40*(1-2), 92-117.

The Comprehensive National Cybersecurity Initiative. The White House. Available at: www.whitehouse.gov

Valeriano, B., and Maness, R. C. (2014). The dynamics of cyber conflict between rival antagonists, 2001–11. *Journal of Peace Research*, *51*(3), 347-360.

Waldrop, M. M. (2016). How to hack the hackers: The human side of cybercrime. *Nature News*, *533*(7602), 164.

Wilner, A.S., 2019. US cyber-deterrence: Practice guiding theory. *Journal of Strategic Studies*, pp.1-36.

Wilson, T., D. Maimon, B. Sobesto, and M. Cukier. 2015. "The Effect of a Surveillance Banner in an Attacked Computer System: Additional Evidence for the Relevance of Restrictive Deterrence in Cyberspace." *Journal of Research in Crime and Delinquency 52*: 829-855.

Willison, R., Lowry, P. B., and Paternoster, R. (2018). A Tale of Two Deterrents: Considering the Role of Absolute and Restrictive Deterrence to Inspire New Directions in Behavioral and Organizational Security Research. *A Tale of Two Deterrents: Considering the Role of Absolute and Restrictive Deterrence in Inspiring New Directions in Behavioral and Organizational Security," Journal of the Association for Information Systems (JAIS)*, *19*(12), 1187-1216.

Workman M, Bommer WH, Straub D. 2008. Security lapses and the omission of information security measures: A threat control model and empirical test. *Computers in Human Behavior* 24:2799-2816.