MUNI FSS

JOURNAL OF PSYCHOSOCIAL RESEARCH ON CYBERSPACE
CYBERPSYCHOLOGY

Pyke A., Rovira, E., Murray, S., Pritts, J., Carp, C. L., & Thomson, R. (2021). Predicting individual differences to cyber attacks: Knowledge, arousal, emotional and trust responses. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 15*(4), Article 9. https://doi.org/10.5817/CP2021-4-9

Predicting Individual Differences to Cyber Attacks: Knowledge, Arousal, Emotional and Trust Responses

Aryn Pyke[1,2], Ericka, Rovira[1], Savannah Murray[1], Joseph Pritts[1], Charlotte L. Carp[3], & Robert Thomson[1,2]

[1] U.S. Military Academy, West Point, New York, USA
[2] Army Cyber Institute, West Point, New York, USA
[3] University of Houston, Houston, Texas, USA

Abstract

Cyber attacks are increasingly commonplace and cause significant disruption, and therefore, have been a focus of much research. The objective of this research was to understand the factors that might lead users to fail to recognize red flags and succumb to cyber events. We investigated users' knowledge of cyber attacks, their propensity to trust technology, arousal, emotional valence, and situational trust in response to different types and severity of cyber attacks. Our findings suggest that high-risk attacks elicited more arousal and more negative emotional valence than low-risk attacks. The attack-type manipulation revealed that phishing scenarios yielded distinctive patterns, including weaker affective responses than ransomware and other malware. The authors further examined arousal, emotional valence, and situational trust patterns among the subset of high- knowledge participants who successfully identified all the attacks and compared these responses with those of less knowledgeable peers. Our findings suggest that the more knowledgeable the user, the higher was their general propensity to trust technology, the more sensitive were their emotional responses to the manipulation of risk, and the lower their situational trust when faced with cyber attack scenarios.

Keywords: Cyber attacks; arousal; trust; phishing; malware; ransomware; emotion; individual differences

Introduction

Networked devices that keep us constantly connected are found in all aspects of our society from work to play, and, as such, heavily influence our daily lives (Baker, 2016). Although this reliance on networked systems has advantages, it can also make us vulnerable to various types of cyber attacks. A cyber attack is an intentional offensive action mounted against our digital devices and/or networks. Such an attack might compromise the functioning of computers or networks and/or result in data being stolen, deleted or altered (Dutt et al., 2013). In the last five years, the cyber attack landscape has been dramatically impacted by the vast amounts of data available as well as the Internet of Things (IoT). IoT allows individuals to connect to networks and access a variety of functions and data via personal devices. IoT capitalizes on the automation of functions previously done by humans, thus allowing for increased communication, increased monitoring of data, as well as huge time and financial savings. Nonetheless, insufficient security awareness and training makes vulnerability to cyber attacks inevitable. The success of a cyber attack is determined by the adversaries' ability to find vulnerabilities and exploit some elements of the system including the human user (Aycock et al., 2008; Pfleeger & Caputo, 2012).

Types of Cyber Attacks and Users' Understanding of Them

Cyber attacks come in a variety of forms and may target organizations or individuals, and cyber attackers may come internally from an organization or externally. Phishing attacks, spyware, trojans, wiper attacks, ransomware, and distributed denial of services are some of the most common forms of cyber attacks. Phishing is an attempt to gain personal information (passwords, credit card information, etc.) usually by soliciting it from users through electronic means like an email which purports to be from a trusted entity (e.g., your bank). Such emails ask users to provide personal information via a reply email or by filling in information on a form or site that appears when they click a link provided in the email. Phishing emails still remain the most common form of cyberattack despite the advent of email filtering systems that try to prevent such emails from reaching user's inboxes (Pricewaterhouse Coopers, 2016). In fact, in a 115-participant study, all participants had experienced a phishing attack (Kelley et al., 2012).

Ransomware is another type of attack that encrypts or prevents access to data on your system. The attacker then demands a ransom payment, often in digital currency (e.g., Bitcoin), to decrypt the data and/or restore access. Randsomware attacks are on the rise. According to a mid-year update (SonicWall, 2021), from 2020 to 2021 there was a 151% rise in randsomware attacks. Other types of malware attacks employ malicious software to delete or steal data, gain backdoor access to the system, monitor user activities and/or open illegitimate popup windows on the user's machine (e.g., viruses, spyware and adware). There were 200,000 new malware samples captured per day in 2016 (Panda Labs, 2016).

In practice, an attack may involve multiple and/or cascading components – so that, for example, a successful phishing attack may secure a user's password, and with that information the cyber attacker can then gain access to the system and install malware programs (e.g., spyware) and/or encrypt files in order to request a ransom. The fact that an attack element need not be used in isolation is very salient and useful from the perspective of a cyber attacker. However, our current focus was on the perspective of the end user. For example, how well could an end user characterize the (current phase/type of an) attack based on the limited salient symptoms evident to them at a given point in time. Such evidence might take the form of an email requesting personal information (phishing), or a message requesting money to restore file access (ransomware), or unwanted popup ads or the mouse moving on its own (malware like Adware or a Remote Access Trojan). We were also interested in end users' affective responses to such events. End users are an important element in the bigger cyber security picture. As emphasized by Kostyuk and Wayne (2021, p. 1), "data breaches with the potential for national, macro-level consequences are most likely to occur at the micro-level, originating through the security errors of individual computer users". Indeed, individual users can be a source of system vulnerability (e.g., getting phished, clicking links that install malware), but they can also facilitate attack detection (e.g., notifying IT of suspicious computer symptoms).

A user's knowledge about cyber attacks and how to handle them is a factor that impacts susceptibility to cyber attacks. This has been demonstrated for phishing attacks. Downs et al. (2007) presented participants with emails that included a link and a URL address. Participants were required to decide how to respond (e.g., reply to email, click on link, or copy and paste the URL). Specific knowledge about phishing and URL addresses predicted lower susceptibility to the emails representing phishing attacks. However general computer and security knowledge was not correlated with participants' likelihood of clicking on a phishing link (Downs et al., 2007). Similarly, Dhamija et al. (2006) found that participants with poor knowledge of computer systems and security features were more likely to incorrectly perceive a fraudulent website as legitimate. On the flip side, improved performance or response to phishing attacks has been demonstrated through the use of a short training intervention (Sheng et al., 2007).

Sawyer and Hancock (2018) suggest that cyber attackers are intentionally taking advantage of the poor ability of humans to detect and recognize rare signals in the environment and are reducing the number of phishing attacks users experience but increasing the impact of the harm. Increasingly researchers are interested in identifying various individual behaviors that may be used to determine who is most and least susceptible to cyber attacks. Yu et al. (2019) recently conducted a study investigating human behavior in response to phishing emails. They found that users that used slow mouse movements had high awareness of phishing emails and that mouse movements could be used to determine users that may fall victim to phishing attacks. Of note however, mouse hovering behaviors were not correlated with phishing awareness.

Trust in Technology

Besides users' knowledge of specific types of attacks, another factor that may influence and predict vulnerability to cyber attacks is trust in technology. Specifically, users may be vulnerable when they inappropriately trust and respond to malicious requests (e.g., to install malware or provide personal information). Given increased automation, the IoT, and the resulting increased diversity of cyber attack susceptibility, it is of interest to understand how trust in technology and cyber security are related. User trust can be considered a human factors issue. As reported by Chong et al. (2019, p. 5) "very few researchers have commented on the value of designing IoT devices with the user as an integral security component". However, the attention towards this issue is growing, and recently Nieto and Rios (2019) provided guidelines to define cybersecurity profiles based on human factors.

Such efforts can be informed by the research investigating trust and automation completed in the last 40 years (for reviews see: Glikson & Woolley, 2020; Hoff & Bashir, 2015; Parasuraman & Wickens, 2008). There are a few prevailing findings: first, when operators are supported by reliable automation technology compared to performing a task manually human performance improves. However, most technological systems are not 100% reliable, and when the automation fails people are caught being too trusting or complacent, resulting in poor performance due to out of the loop unfamiliarity or a loss of situation awareness (Parasuraman & Wickens, 2008). Increasingly researchers are interested in understanding individual differences in automation bias (J. Y. C. Chen & Terrence, 2009; de Visser et al., 2010, Parasuraman et al., 2012; Rovira et al., 2017), where automation bias is the propensity to trust in decision-making by automated systems and ignore contradictory information even if it is correct (Mosier et al., 1998).

Consequently, recent efforts in the automation literature are focused on trust calibration, which refers to individuals calibrating their trust appropriately to a specific system or piece of technology. For cybersecurity, a relevant application will be the calibration of user trust in automated security tools that provide recommendations such as whether a website is safe or malicious, but allow users to make the final security decisions (e.g., Y. Chen et al., 2021). Currently, however, most users do not have a user-friendly all-purpose automated security tool that can reliably advise them about which apps, emails, websites, pop-ups etc. are safe. Thus currently, it is not a users' trust in a specific automated security application, but their trust in technology in general that likely predicts their propensity to comply with, say, instructions in a pop-up window that appears on their screen. In this context, too much "trust in technology" can be risky.

Individuals overtrust technology for a variety of reasons. The temporal or mental demand of the task may be too high, they may have positive experiences with the technology working reliably most of the time, or the cost of the technology failing is not catastrophic. Increasing our understanding of why some individuals succumb to cyber attacks and why others do not may very well be linked to individual differences in propensity to trust technology.

There is a distinction between a general propensity to trust technology and the situational trust of a user in response to a particular scenario. In the extreme, a general lack of trust in technology may result in avoiding the use of technology (Kelley et al., 2012), and forgoing the benefits in terms of complexity and speed that arise from automation (Lee & See, 2004). However, the reduction of trust in response to a current situation (situational trust) is presumably transient. Kelley et al. (2012) found that falling prey to a phishing attack resulted in reduced trust in technology (and other people) and an increased unwillingness to use a service in the future.

Affect: Arousal and Emotional Valence

Computer interaction can impact user affect, and user affect can in turn impact how users interact with computers. For example, pop-up messages which suggest that the user make software updates can produce emotional responses like anxiety, sadness, annoyance, or confusion (Buck et al., 2017). The level of emotion experienced has been demonstrated to predict users' hesitation in applying software updates - e.g., the more annoyed, the less compliant (Fagan et al., 2015). In the current research we are interested in investigating users' affective responses to cyber attacks (phishing, malware and ransomware).

Some prior work has shed light on affective responses to cyber attacks more generally. Cyber harm is a broad term encompassing harm of various types (e.g., psychological, physical, economic, social, etc.) perpetrated via the

Internet or other electronic means typically involving a cyber-incident or attack (Agrafiotis et al., 2018). Within the cyber-harm taxonomy of Agrafiotis et al. (2018), aspects of psychological harm include, among others, feeling confused, upset, frustrated, anxious, depressed, and embarrassed. These and other affective responses were reportedly experienced by employees and customers of companies, like Sony, that were impacted by a cyber attack, and by victims of identity theft (Hess, 2015). There are also some experimental studies that assess participants' responses to hypothetical cyber attacks that could be classified as "cyber terrorism". For example, Gross et al. (2016; see also 2017) investigated Israeli participants' responses to four hypothetical scenarios (control with no attack, non-lethal cyber terrorism, lethal cyber terrorism and lethal conventional terrorism). They assessed participants' anxiety levels (State-Trait Anxiety Inventory: STAI) and political reactions (type and scale of retaliation favored, willingness to forgo privacy to allow preventative surveillance). Reported levels of anxiety were lowest for the control scenario, highest for conventional terrorism, and intermediate and comparable for lethal and non-lethal cyber terrorism scenarios. In another study, researchers found that relative to a no-attack control group, participants exposed to a hypothetical cyber attack scenario (non-physical/non-lethal) exhibited an increase in the stress hormone cortisol in their saliva (Canetti et al., 2017). Critical infrastructure systems are desirable targets for cyber terrorism that can yeild emotional responses (Backhaus et al., 2020).

In the current research, however, we were interested in affective responses to the types of cyber attacks that might be more commonly experienced by an individual computer user, such as phishing and malware. Understanding the relative affective responses to different types of attacks that individuals experience might foster better tailored strategies for mitigating risk. Strong affective responses have been shown to decrease performance and cause errors in judgement (Cohen, 2011). Thus, a user's affective reaction may predict how well the user is able to act in a way that minimizes damage and/or precludes falling prey to the attack (i.e., avoid clicking on a suspicious link). Consequently, we were also interested in factors that might impact individual differences in these affective responses, including trust in automation and cyber knowledge. Canetti et al. (2017) did not find a correlation between their computer knowledge measure and their participants' stress response (cortisol) to attack scenarios, however their computer knowledge measure involved self-reported aptitude for relatively common tasks (e.g., email, word processing, using a search engine). Another study assessed cyber knowledge in terms of knowledge of cyber events in the news ("cyber event awareness", Gomez, 2019), but did not include a direct measure of stress or affect to test for a knowledge-affect relation. Our knowledge measure involved the ability to identify the particular types of cyber attack occurring in each stimulus scenario.

As detailed more fully in the Methods section, to assess affective response, our participants were asked to rate both their emotional valence (sad to happy) and arousal levels (low to high) in response to our hypothetical cyber scenarios using 5-category pictorial Self-Assessment Manikin scales (Bradley & Lang, 1994). The distinction between emotion and arousal is not exact in the literature, however some frameworks regard them as different dimensions of affect (Bradley & Lang, 1994). Thus, we included both measures in case they exhibited distinct relations with other factors.

Present Research: Questions and Hypotheses

The current research sought to fill in the gaps in the literature that influence responses to cyber attacks by investigating:

1. The extent to which the type and severity of a cyber attack impacts trust, arousal, and emotional valence.
2. The extent to which cyber knowledge and propensity to trust technology moderates trust, arousal, and emotional valence responses to an attack.
3. Are there different response patterns between those who can identify cyber attacks and those who cannot?

We hypothesized that high-risk attacks would elicit less trust, more arousal, and more negative emotional valence than low-risk attacks. The attack type manipulation was more exploratory. Since many studies only include a single type of attack (e.g., phishing) (Downs et al., 2007; Kelley et al., 2012), we wanted to assess whether trust and affective response patterns tend to generalize across attack types, and, if not, to provide a broader understanding of user response patterns. We had some intuitive reasons to expect potentially different levels of affective response to phishing versus ransomware versus [other] malware. When a user becomes aware of a malware or ransomware attack, the damage has usually already been done. A phishing attack, however, requires the user's

voluntary participation to provide personal information, which can be avoided if the email or message is recognized as a phishing attempt. Given this higher degree of user control, we suspected that phishing attacks might invoke the most muted emotional responses. In contrast we suspected that a ransomware attack might invoke the strongest responses relative to phishing or other malware attacks because: i) a request for money might make the attack feel more like being the victim of a physical crime like robbery (vs. someone just messing with your computer), and being asked for ransom could also invoke associations with other physical threat concepts like kidnapping and being held hostage; ii) being explicitly asked for money by an attacker also adds insult to injury (i.e., file inaccessibility), and emphasizes the human agency in the attack process, which may make users take these attacks more personally than other malware attacks where what is salient is that your computer is acting funny rather than that another human is attacking you.

We also hypothesized that pre-experimental trust in technology and knowledge about types of cyber attacks (which users presumably possessed pre-experimentally) would be predictors of arousal, emotional valence and situational trust responses to the cyber scenarios. Specifically, we hypothesized that having a strong general tendency to trust technology might make one more vulnerable to cyber attacks and prone to overly high situational trust (i.e., less likely to recognize a red flag). This relation was expected because our pre-experimental trust measure (Merritt et al., 2013) is a measure of "propensity to trust", which is regarded as a relatively stable trait reflecting a general willingness to trust (Chughtai & Buckley, 2008; Koelsch et al., 2008; Mayer et al., 1995; Mooradian et al., 2006).

In contrast, we predicted that those with greater knowledge would report less situational trust, because they would be well aware of the attacks at hand and would adjust their situational trust accordingly. We especially aimed to predict situational trust because this judgment would presumably have the most direct influence on the user's subsequent decisions about what actions to take. For example, if one's situational trust is high, one is presumably more likely to comply with (malicious) requests to execute a piece of malware code or provide personal information. This relationship between trust and compliance has been evidenced in other domains - for example, level of trust in medical organizations predicts vaccination compliance behavior (Gilles et al., 2011; Prati et al., 2011). Further, since it is known that unexpected (vs. expected) events elicit emotional responses (Koelsch et al., 2008), we expected that greater cyber attack knowledge would predict more muted emotional and arousal responses, because such knowledge might make the cyber scenarios less unexpected and/or shocking. We further expected that the arousal, emotion and situational trust responses would all be related, however, we also expected that the relation between arousal and situational trust might be mediated by emotional valence. This expectation was motivated by the reasoning that the information content across these measures might be sequentially supplemented (increased) if users progressed from arousal to emotional valence to situational trust. Specifically, we assumed that arousal was the most visceral and automatic affective response. The emotional response would presumably be influenced by arousal but would also include some kind of cognitive attribution to determine valence (happy or sad). Finally, the situational trust response would be affected by emotion (and, by association, arousal) but would also include a judgement about the wisdom of one's compliance in the situation (e.g., to click the link and/or execute the code).

Finally, in the context of our manipulations, we also hoped to characterize the response patterns (arousal, emotional valence and situational trust) of high knowledge participants, who could correctly identify the attacks at hand, and contrast these response patterns with those of lower-knowledge participants.

Method

Participants

An a priori power analysis using Faul and Erdfelder's (1992) Gpower showed that a minimum of 132 participants were required for enough power to detect an effect size of 0.2 (a power level of .8 and alpha at .05). A total of 142 American college students (32 women) participated in this study for extra credit in a psychology course that is taken by all students (i.e., regardless of academic major) as a general education requirement at this institution. The population had a mean age of 19.5 years (*SD* = 1.4 years). This study was approved by the institution's IRB and data were collected between 2017–2018.

Materials

The study was carried out remotely using Qualtrics, a web-based survey platform. Prior to the factorial survey, participants completed a short propensity to trust survey (Merritt et al., 2013).

Propensity to Trust Survey

The propensity to trust survey consisted of six questions assessing participants overall trust and reliance on technology (Merritt et al., 2013): for example, "It is easy for me to trust technology to do its job." The full list of questions is provided in Table 1. Responses to questions were measured on a five-point Likert scale ranging from 1 = *strongly disagree* to 5 = *strongly agree*. Merritt et al. (2013) reported an alpha of .87 for this set of items. This established measure of an individual's propensity to trust technology is general. In the current context involving a college population, the most prevalent technological devices are mobile phones, laptops, tablets, GPS watches, or cars).

Table 1. *Pre-Experimental Propensity to Trust in Technology Survey Results*.

Questions	***M*[a]**	***SD***
Q1. I usually trust technology until there is a reason not to.	3.8	0.87
(Q2. For the most part, I distrust technology.)	2.2	0.96
Q2. Reverse-coded (as if it said: For the most part, I trust technology)	3.8	0.96
Q3. In general, I would rely on a technology to assist me.	4.0	0.75
Q4. My tendency to trust technology is high.	3.8	0.83
Q5. It is easy for me to trust technology to do its job.	3.9	0.75
Q6. I am likely to trust technology even when I have little knowledge about it.	3.3	1.02

Note. [a] Responses were made using a five category Likert scale ranging from strongly disagree (1) to strongly agree (5).

Cyber Attack Scenarios

Factorial surveys are used to gather subjective assessments by using scenarios. Factorial surveys are beneficial in measuring how experimental manipulations impact subjective perceptions; e.g., trust (Rossi & Anderson, 1982). Factorial surveys have been used in human factors and automation research (e.g., Endsley & Kiris, 1995; Mosier et al., 2012; Pak et al., 2014; Rovira et al., 2019).

The scenarios manipulated three of the most common types of cyber attacks (phishing, malware, ransomware) experienced by individuals and the level of risk posed by the cyber event (low, high). All three types of attacks included in the study (phishing, ransomware and [other] malware) are prolific in cyberspace, so user reactions to each are highly relevant. Phishing is an attempt to gain user information (passwords, credit card information, etc.) usually through electronic means (e.g., email). Malware is malicious software that may be used, for example, to delete or steal data, to gain backdoor access to the system, and/or monitor user activities (e.g., viruses and spyware). Ransomware encrypts or prevents access to data on your system, and the attacker demands a ransom payment, usually in digital currency (e.g., Bitcoin), to decrypt the data and/or restore access.

Participants did not in fact experience an actual cyber event but instead were asked to think about the specific scenarios as if they were experiencing them personally. As such, all vignettes began with the phrase "You are" followed by a description of a situation we wanted them to imagine being in (e.g., reading an email asking for password information, seeing the cursor scrolling through sensitive documents on its own, or having all files related to a big event you've been planning locked via ransomware). Risk level of the situation was operationalized as the severity of the risk posed by the cyber attack. More generally, risk is a construct that can refer to a combination of outcome likelihood and outcome severity (e.g., Wogalter et al., 1999). In the current study, the focus was on the severity dimension of risk, and likelihood was controlled in the sense that the scenarios described a cyber attack that was already occurring (likelihood = 1). In this context, low risk was classified as generic documents not requiring security clearance or involving small sums of money. High risk was classified as documents requiring security clearance, involving large sums of money and/or impacting more people (e.g., the event planning ransomware scenario). The factorial combinations of the two manipulated factors (3 attack types

× 2 risk levels) resulted in 6 unique scenarios. A sample scenario representing a *phishing* cyber attack and a *high risk* situation is represented below:

> *You are a cadet[1] who invests money in the stock market. You* ***receive an email*** *saying that several of your bank accounts are locked because of failed password attempts. In order to get into your accounts,* ***you must submit personal information****. You are in a rut* ***because a lot of money*** *is on the line.*

Key sections that identify this scenario as a phishing cyber attack and high risk are bolded for illustrative purposes. All scenarios are presented in Table 2. The presentation order of the scenarios was randomized for each participant so that the attack type and risk level did not follow a predictable pattern. This condition-order randomization also served to control for any potential systematic carryover effects across scenarios.

Table 2. *Cyber Attack Scenarios That Varied by Attack Type and Risk Level*.

	Type of Cyber attack		
	Ransomware	**Malware**	**Phishing**
High Risk	You are the S3[a] for a cadet[b] club using a computer and realize all of the files are locked. A pop-up displayed prompts you to pay $2,000,000 to access any files. You realize that all the files from a club event you are planning could be lost if the files are permanently locked.	You are a cadet using a government laptop when you notice the pointer moving by itself. You see the pointer click on and scroll through secret information. You are not sure what is going on, but it seems as if the computer has a mind of its own.	You are a cadet who invests money in the stock market. You receive an email saying that several of your bank accounts are locked because of failed password attempts. In order to get into your accounts, you must submit personal information. You are in a rut because a lot of money is on the line.
Low Risk	You are a cadet, streaming Netflix on a laptop when it freezes and displays a pop-up. The pop-up says that the computer will remain frozen unless you pay a $1 ransom.	You are a cadet downloading a video game from the internet. After you click install, several ads pop up on your screen. You try to close them, but more continue to pop up.	You are a cadet who received an email from Apple. The email prompts you to select a link, and input your Apple ID and password. The email explains that if you do not enter your credentials, all of your iTunes content will be deleted.

Note. [a] The S3 role is the operations officer who, among other things, plans events. [b] Concurrently with their university education, the students in our sample also receive training/instruction to enable them to serve as Army Officers upon graduation. In that capacity they can be referred to as cadets, so the use of the term cadet in the vignettes was intended to make them relatable to these participants.
Flesch Kincaid Reading Ease = 80.3.
Flesch Kincaid Grade Level = 4.8.

Flesch–Kincaid readability statistics (Kincaid et al., 1975) showed that the mean grade level for the scenarios was 6.35. All scenarios were pilot tested to ensure that each manipulation (type of cyber event and risk) was noticeable. In the pilot test, undergraduate participants read each scenario and were asked to identify the type of cyber attack and judge whether it seemed to be a low or high risk. Pilot participants were accurately able to state the type of cyber event and relative risk (low or high) as intended.

Experimental Design

The experimental design was a 3 (type of cyber attack: phishing, malware, ransomware) × 2 (risk level: low, high) within-subjects design with each participant exposed to every type of cyber event and risk level. Participants were each presented with six cyber event scenarios that were constructed by crossing two independent variables: type of cyber attack (ransomware, malware, or phishing) and risk level (low vs. high). The 6 scenarios were presented in a random order for each participant one-at-a-time.

Dependent Variables

Four questions were posed to participants in response to each scenario measuring: situational trust in technology, arousal, emotional valence, and a knowledge measure (identify the type of attack). The first question on situational trust measured how much participants "trust the technology in the particular scenario" (Lee & Moray, 1994). Responses were assessed on a seven category Likert scale ranging from 1 = *strongly disagree* to 7 = *strongly agree*. For analysis, these responses were re-coded to the range 1 = *strongly disagree* to 5 = *strongly agree* for consistency

with the pre-experimental trust responses. Recoding preserved the equal interval property of the scale: 1 to 1, 2 to 1.6667, 3 to 2.3333, 4 to 3, 5 to 3.6667, 6 to 4.3333, 7 to 5. Thus, higher values reflect more trust, with 5 as the maximum.

The next two questions required participants to rate their emotional valence and arousal levels in response to the scenario using 5-category pictorial Self-Assessment Manikin scales (Bradley & Lang, 1994). For emotional valence, the manikins depicted on the scale ranged from a frowning, unhappy figure to a happy smiling one. For analysis, emotional valence responses were coded on a 5 point scale with -2 = *very sad*, 0 = *neutral* and 2 = *very happy*. Thus, higher values reflect a more positive emotional response. On the arousal scale, the manikins ranged from a relaxed sleeping figure to a wide-eyed excited one. Arousal responses were numerically coded on a 5-point scale (1 = *very low* to 5 = *very high*). Thus, higher numerical values reflect higher levels of arousal.

The final question measured the participant's knowledge about the type of attack described in the scenario. Using a multiple-choice response, participants were able to select via a drop down: malware, ransomware, phishing, or other. If participants selected other, then there was a blank text box in which participants could type in a response. We sought to assess whether users were equally able to identify all three types of attacks. We acknowledge that malware can be used as an umbrella term for various types of cyber attacks to include phishing and ransomware. In retrospect we could have perhaps refined the wording of the choices to be: phishing, ransomware, **other malware**, and other. However, as per conventional multiple-choice protocol, we expected participants to choose the best answer. For example, in the case of a ransomware scenario, given the available options (phishing, ransomware, malware and other), the arguably best (i.e., most precise) answer would be ransomware. To foreshadow, participants' attack type responses suggest they seemed to follow this protocol.

Procedure

After signing up to participate via Sona, participants were sent a link to complete the survey on Qualtrics. Sona is a website that allows researchers to post research studies and allows participants to sign up for research. Participants were given instructions to complete the survey in one sitting and to avoid taking breaks (web log time entries confirmed this). Participants completed the experiment remotely and were instructed to use their own laptops. Prior to judging all 6 scenarios, participants completed the pre-experimental propensity to trust in technology survey.

Participants were then presented with one of the six scenarios. After reading the scenario, participants were required to answer the four questions (i.e., about their situational trust, emotional valence, arousal, and knowledge of attack type). After a participant had responded to all four questions, another scenario was presented followed by the same four questions. This procedure continued until participants had been exposed to all six scenarios. Lastly, participants provided some demographic information (i.e., age and gender). Participants were then thanked for their participation and awarded their extra credit points. The study took approximately 20 minutes to complete.

Results

We first present the results of the pre-experimental trust questionnaire, and then the experimental results are structured into two sections corresponding to our two agendas for these data: i) analyses in ***Effects of Manipulations: Attack Type × Risk Level*** investigate the impact of our manipulations of attack type and risk level; and ii) analyses of ***Individual Differences: Predicting Affective Responses*** collapse across attack type to investigate and model general relations among response measures and between these measures and the predictors of pre-experimental trust and cyber attack knowledge.

Pre-Experimental Trust in Technology

Recall that participants responded to questions about their propensity to trust technology using a scale from 1 = *strongly disagree* to 5 = *strongly agree*. As evident in Table 1, five of the six questions were worded in a pro-trust manner (e.g., "In general, I would rely on a machine to assist me", Merritt et al., 2013). The question worded in an anti-trust manner (Q2. "For the most part I distrust machines") was reverse-coded to be consistent with the other

questions (i.e., so 5 would reflect maximum trust in technology). Cronbach's alpha was calculated to be .834, which suggests that the items had relatively high internal consistency. Table 1 presents the means and standard deviations of the responses for each question. The overall mean (M = 3.77, SD = 0.64) suggested that our participants had a tendency to trust technology, but that this tendency was not a very strong one - i.e., the mean agreement with statements indicating trust in technology was above "neutral" (3) but below "agree" (4). Some tendency to trust technology is unsurprising given that our sample consisted of college students who could be considered digital natives (mean age 19.5 years). Our sample also further evidenced comfort with technology by volunteering to participate in study involving an on-line survey. However, as this college student population uses computers and the internet daily, they also have first-hand familiarity with the glitches associated with such technology. We will revisit these data in the individual differences section when exploring relations between pre-experimental trust and experimental response patterns. In that context, users' pre-experimental trust will be indexed as the mean of responses across all six survey questions (with Q2 reverse-coded).

Effects of Manipulations: Attack Type × Risk Level

In this section we report results about the impact of our manipulations of attack type and risk level on the four response measures to the scenarios: arousal, emotional valence, situational trust and knowledge (identification of the type of attack). For each of these dependent measures, we conducted a 3(attack type: ransomware, malware, phishing) × 2(risk level: high, low) repeated-measures ANOVA[2]. The means and standard deviations of the dependent measures are summarized in Table 3 for each of the six conditions. In the ANOVA analyses, if the 3-level attack-type factor violated Mauchly's Test of Sphericity, the Greenhouse-Geisser correction was used to assess significance. Finally, when doing pairwise comparisons to explore main effects of the 3-level attack-type factor, these pairwise comparisons were made using a Bonferroni adjustment. Reported p values for these comparisons are copied from the SPSS output and reflect multiplying the uncorrected p by the number of comparisons and considering it significant if it is below .05. This yields the same conclusions as comparing the uncorrected p value to a threshold of (.05/number-of-comparisons).

Table 3. *Summary of Emotion, Arousal, Situational Trust and Knowledge Results*.

Cyber Scenario	Emotion -2=very sad +2=very happy		Arousal 1=min, 5=max		Situational Trust 1=min, 5=max		Knowledge % Accuracy	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
All participants (*N* = 142)								
Ransomware								
Low Risk	-0.68	0.86	3.02	1.00	1.50	0.67	75.4	43.2
High Risk	-1.06	0.87	3.34	1.25	1.48	0.72	66.9	47.2
Malware								
Low Risk	-0.67	0.73	2.96	0.96	1.78	0.80	85.2	35.6
High Risk	-1.18	0.95	3.46	1.36	1.44	0.82	70.4	45.8
Phishing								
Low Risk	-0.34	0.80	2.80	1.05	1.85	0.96	78.2	41.5
High Risk	-0.64	0.87	3.16	1.10	2.06	1.10	75.4	43.2
Participants Who Correctly Identified All attacks (*N* = 47)								
Ransomware								
Low Risk	-0.72	0.88	2.91	1.02	1.48	0.63	100	0
High Risk	-1.23	0.87	3.15	1.41	1.50	0.79	100	0
Malware								
Low Risk	-0.66	0.76	2.98	1.03	1.68	0.62	100	0
High Risk	-1.45	0.80	3.57	1.47	1.34	0.64	100	0
Phishing								
Low Risk	-0.15	0.75	2.77	1.22	1.72	0.89	100	0
High Risk	-0.49	0.91	3.02	1.05	1.99	1.09	100	0

Arousal

The arousal results are presented in Figure 1a. As expected, participants reported higher levels of arousal for high-risk (M = 3.3, SE = 0.08) versus low-risk attack scenarios (M = 2.9, SE = 0.06), $F(1,139)$ = 23.61, partial η^2 = 0.15,

$p < .001$[3]. Mauchly's test indicated that the assumption of sphericity was violated for attack type ($\chi^2(2) = 9.01$, $p = .011$), so degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.94$). This yielded a main effect of attack type, $F(1.91, 260.87) = 4.76$, $p = .011$[4]. Specifically, arousal was lower for phishing attacks ($M = 3.0$, $SE = 0.07$) than for both ransomware ($M = 3.2$, $SE = 0.08$, $p = .042$) and malware attacks ($M = 3.2$, $SE = 0.08$, $p = .026$), which did not significantly differ ($p > .999$). The interaction was not significant.

Emotional Valence

The pattern of emotional responses is presented in Figure 1b. Participants reported more negative emotion for high-risk ($M = -0.96$, $SE = 0.06$) versus low-risk scenarios ($M = -0.56$, $SE = 0.05$), $F(1,139) = 58.69$, partial $\eta^2 = 0.30$, $p < .001$[5]. Mauchly's test indicated that the assumption of sphericity had been violated for the attack type factor ($\chi^2(2) = 8.04$, $p = .018$), so degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.95$). This revealed a main effect of attack type on emotional response, $F(1.89, 263.10) = 27.52$, partial $\eta^2 = 0.17$, $p < .001$. In comparison to phishing attacks ($M = -0.49$, $SE = 0.06$), more negative emotional responses were reported in response to ransomware ($M = -0.87$, $SE = 0.06$, $p < .001$) and malware attacks ($M = -0.93$, $SE = 0.06$, $p < .001$), which did not significantly differ ($p > .999$). Finally, the interaction was not significant. In this within-subjects design, participants saw all the scenarios. The condition-order randomization across subjects served to control for any potential *systematic* carryover effects across scenarios, however due to exposure to multiple scenarios, the mean emotion valence may be lower and the mean arousal ratings may be higher in this study than in a design where subjects only saw one of these scenarios. As evident from Table 3, however, overall responses were fairly moderate, which might be expected as the scenarios were merely hypothetical.

Situational Trust in Technology

Trust ratings are summarized in Figure 1c. To provide a sense of how these situational trust levels compared to participants' pre-experimental trust in technology, the mean pre-experimental trust level among questions (with Q2 reverse coded) is provided as a reference line in Figure 1c. As evident from Figure 1c, when faced with a specific cyber attack, a participant's situational trust in technology ($M = 1.70$, $SD = 0.64$) falls well below their general (pre-experimental) trust level ($M = 3.77$, $SD = 0.57$), $t(141) = 28.36$, $p < .001$. We acknowledge, however, that pre-experimental and situational trust measures involved different questions which may raise some concerns about the direct comparison above. The analysis that follows is based solely on the situational trust measure.

Mauchly's test indicated that the assumption of sphericity was violated for attack type ($\chi^2(2) = 19.21$, $p < .001$), so degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.88$). This yielded a main effect of attack type on participants' situational trust in technology, $F(1.76, 236.23) = 26.98$, partial $\eta^2 = 0.17$, $p<.001$[6]. Technological trust remained higher during phishing attacks than malware or ransomware attacks ($ps<.001$). Malware attacks received marginally higher trust ratings than ransomware attacks ($p = .085$). There was no main effect of risk level on participants' situational trust. For the risk level by attack type interaction, again Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(2) = 9.80$, $p = .007$), so degrees of freedom were corrected via Greenhouse-Geisser estimates ($\varepsilon = 0.93$) and yielded an attack type by risk level interaction, $F(1.86, 250.23) = 12.08$, partial $\eta^2 = 0.07$, $p < .001$. For ransomware attacks, the resulting trust in technology did not significantly differ between low- and high-risk conditions ($p = .714$), but for malware, trust decreased as risk increased ($p<.001$), and for phishing attacks, reported trust increased as risk increased ($p = .042$).

To summarize the results so far, arousal and emotional valence were both sensitive to the risk-level manipulation – higher risk corresponded to more arousal and a more negative emotional response. Further, in terms of attack type, phishing attacks always yielded the least negative affective responses (least arousal, less negative emotional valence) and the highest situational trust.

Knowledge: Identification of Attack Type

Preliminary examination of the data revealed that when specifying the type of attack in each scenario, participants typically selected one of three specific options (malware, ransomware, phishing) as opposed to the "other" option (typically < 5% of responses). An exception to this behavior was the frequent choice of the "other" category for the malware high-risk scenario (28% of responses). Participants were prompted to specify the alternative attack type,

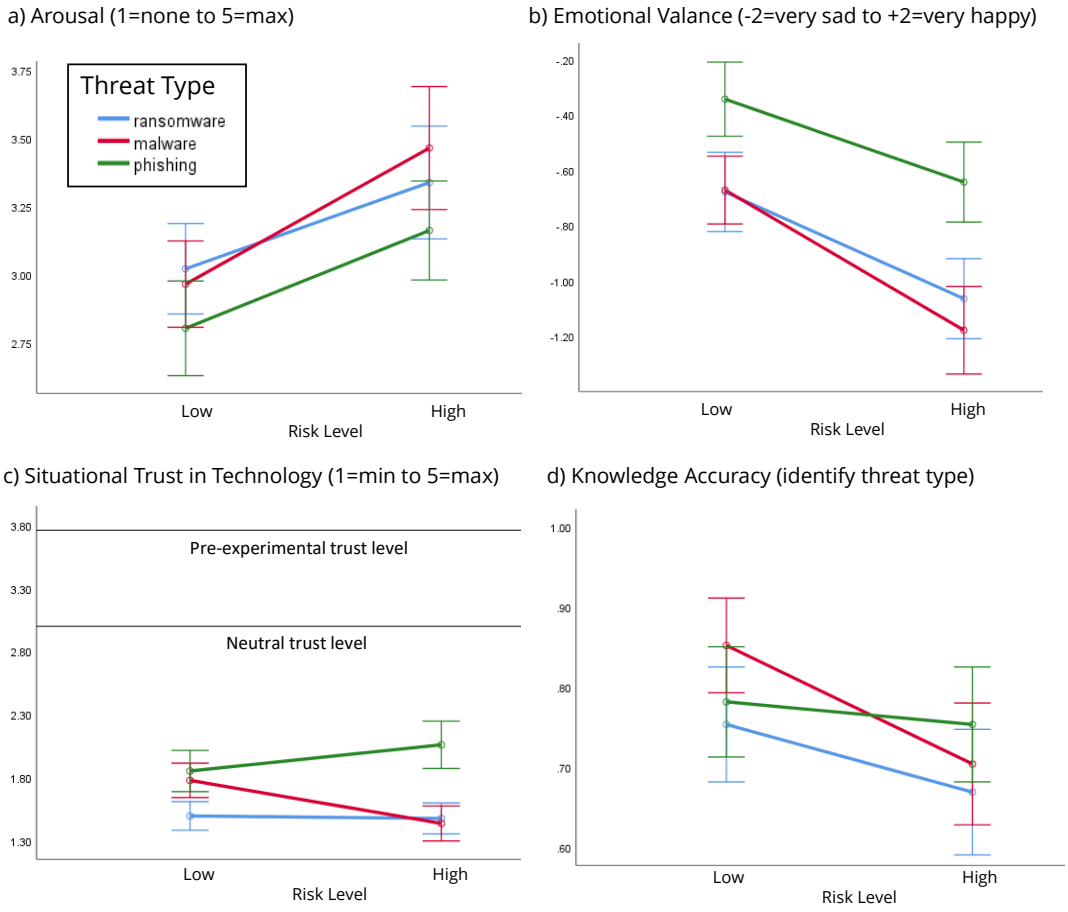
and several of these responses proved to be correct possibilities for the malware scenarios (e.g., remote access and remote hacking for Malware High). Table 4 provides a breakdown of attack-type responses and indicates (with bold numbers) the "other" answers that were scored as correct in the following analyses of knowledge accuracy (and in Table 3).

Table 4. *Cyber Knowledge: Percent Breakdown of Attack Type Responses to the Scenarios*.

	Scenario					
Participant's Attack Label	Ransomware Low	Ransomware High	Phishing Low	Phishing High	Malware Low	Malware High
Ransomware	**75.4%**[a]	**66.9%**	9.9%	12.7%	5.6%	9.9%
Phishing	14.8%	13.4%	**78.2%**	**75.4%**	6.3%	7.0%
Malware	8.5%	18.3%	9.9%	8.5%	**81.7%**	**54.9%**
Other	1.4%	1.4%	2.1%	3.5%	6.3%	28.2%
[blank/missing]						2.1%
Don't know[b]	1.4%	0.7%			0.7%	7.0%
Not an Attack[c]		0.7%	0.7%	2.8%		2.1%
Cyber Attack						0.7%
Remote Access[d]						**15.4%**
Trojan						**0.7%**
Spam					0.7%	
Adware					3.5%	
Advertising/Popups					1.4%	

Note. [a] Bold percent values were scored as correct. [b] includes: [I] don't know, unknown, [I'm] not sure, no idea, not quite sure, etc. [c] includes: none, nothing, not an attack, It wasn't an attack, computer malfunction, technical difficulties. [d] includes: remote hack[ing], remote login attack, hack/hacked/hacking, hacking: remote control, etc.

Figure 1. *Means of Experimental Responses by Attack Type and Risk Level for: a) Arousal, b) Emotional Valence, c) Situational Trust in Technology and d) Knowledge Accuracy at Identifying the Attack Type*.



Note. Error bars are 95% confidence intervals.

Participants' overall accuracy at identifying the types of cyber attacks described in the scenarios was 75%. Figure 1d summarizes attack identification accuracy by condition. Participants were less accurate at identifying the type of attack in high than low risk scenarios (M = 71%, SE = 2.6 vs. M = 80%, SE = 2.2, p = .001), $F(1, 141) = 11.58$, partial η^2 = .08, p = .001. Neither the effect of attack type, nor the interaction between attack type and risk level reached significance.

Response Patterns of High Versus Low Knowledge Participants

About a third of the participants (47/142) correctly identified the type of attack in all six scenarios. Since it may be a useful contribution to characterize the response profiles (arousal, emotion and situational trust) of users who are savvy about these attacks, we have provided the response data for this subset of users in Table 3. We also performed supplementary analyses to compare their arousal, emotion and situational trust responses with those of less cyber-savvy peers at the other end of the spectrum (i.e., low-knowledge participants who got < 50% correct, N = 32). Specifically, we sought to detect any effects or interactions of knowledge level in these supplementary 3(attack type) × 2(risk level) × 2 (knowledge level: < 50% vs. 100%) ANOVAs for each of the response measures.

For arousal, although the perfect-score participants had numerically lower arousal levels (M = 3.01, SE = 0.11) than their low-knowledge participants (M = 3.3, SE = 0.13), neither this difference nor any of the knowledge interactions reached significance. For situational trust, levels were marginally lower among participants with perfect scores (M = 1.6, SE = 0.09) than low-knowledge participants (M = 1.84; SE = 0.10), $F(1,74) = 3.00$, partial η^2 = .39, p = .087. No interactions with knowledge level reached significance for situational trust.

For emotional valence, there were interactions of knowledge level with both risk level, $F(1,77) = 6.54$, partial η^2 = .08, p = .013, and attack type, $F(2,154) = 6.88$, partial η^2 = .08, p = .001. Specifically, perfect-score participants had emotional responses that were more distinctive across high versus low-risk scenarios (High M = -1.06, SE = 0.10 vs. Low M = -0.51, SE = 0.08) than did low-knowledge participants (High M = -0.89, SE = 0.12 vs. Low M = -0.69, SE = 0.10). That is, knowledgeable participants exhibited more sensitivity to the risk manipulation in their emotional responses. In terms of attack type, for phishing scenarios, high-knowledge participants had less negative emotional responses (M = -0.32, SE = 0.10) than their low-knowledge peers (M = -0.70, SE = 0.12, p = .014). Responses did not differ significantly for the other attack types.

High knowledge participants could consistently identify/classify the type of attack at hand. Subjects who did not correctly identify all the cyber attacks in the scenarios, however, are arguably a less homogeneous set. Being unable to correctly identify the attack did not necessarily mean they were oblivious to it, since they often misidentified the attack at hand as another type of attack. In general, we expected that the measure of situational trust might provide a more nuanced measure of a participant's level of attack awareness. Thus, rather than making categorical comparisons (e.g., across high and low knowledge participants as we did in the above supplementary analyses) in the individual differences section below we explore relations across the response spectrum.

Individual Differences

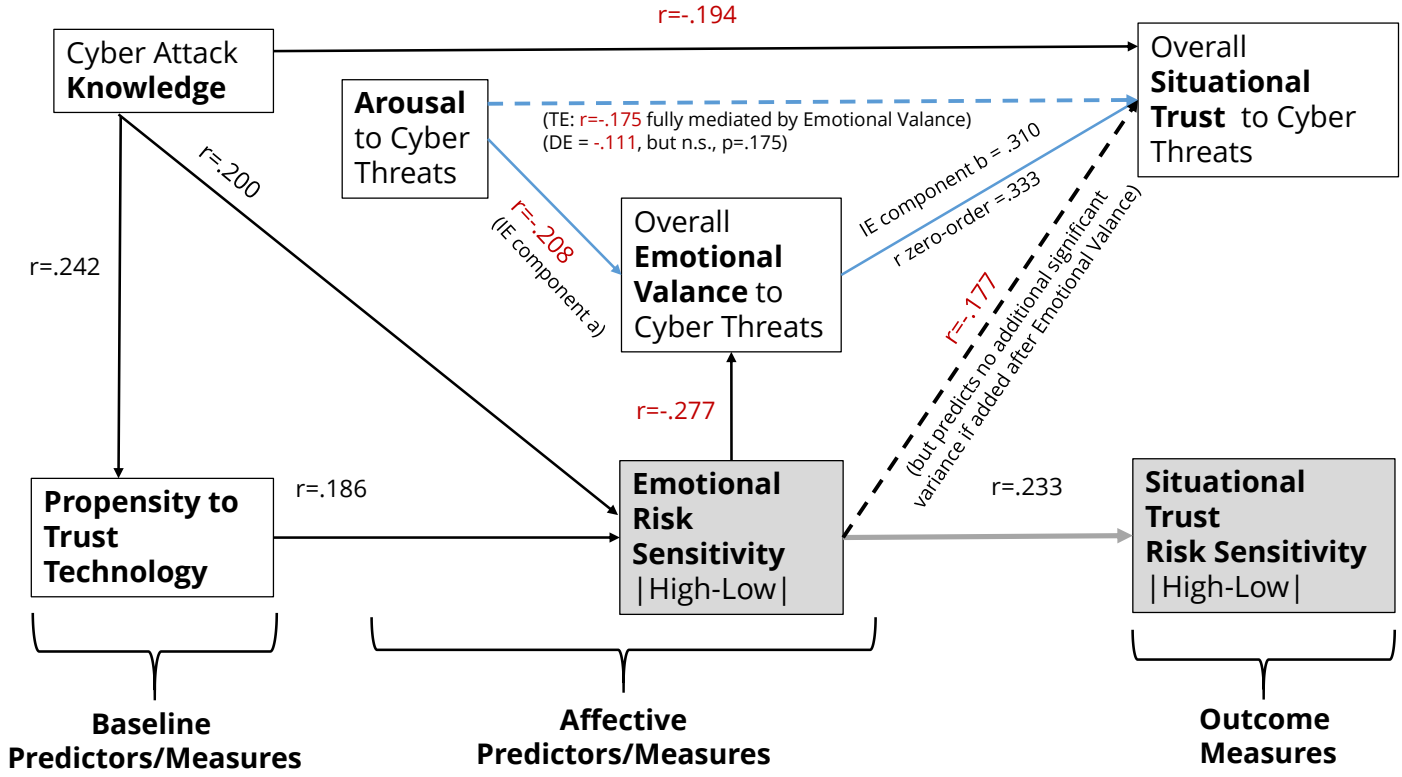
Our objective in this subsection of results was to abstract across attack type to develop a model to understand relations among experimental responses and, if possible, predict them using pre-experimental trust and cyber attack knowledge. However, based on the above analyses (see Figure 1), while malware and ransomware responses seemed closely coupled, phishing responses tended to significantly differ from the other two, and also yielded an unexpected increase in situational trust with risk level. Further, preliminary correlation analyses revealed that increased attack knowledge predicted more positive emotional responses for phishing (r = .198, p = .018, 2-tailed) whereas the trend was for more negative emotional responses for malware and ransomware (r = -.110 and r = -.128, *ns*). In light of these differences, the following individual differences model is based on affective responses for malware and ransomware. For phishing, of the four other main measures, only emotional valence significantly predicted situational trust (r = .263, p = .002, 2-tailed). Emotional valence was also related to knowledge (r = .198, p = .018, 2-tailed) and arousal (r = -.236, p = .005, 2-tailed). Correlations reported in this section are 2-tailed.

As discussed in the Present Research section, we were particularly interested in a model for predicting situational trust (i.e., the degree to which the cyber-attack scenarios raised red flags for the subject). The situational trust measure was of particular interest because it will arguably have the most direct influence on the user's subsequent decision about what action to take (or not to take). We actually considered two facets of situational trust to be predicted: i) an overall situational trust level in response to the cyber attacks, collapsed across attack type (malware and ransomware) and risk level (high, low); and ii) degree of risk sensitivity in situational trust (|high-low|)[7]. Note that these separate facets of situational trust (overall vs. risk-sensitivity) were not correlated (r = -.010, p = .905, *ns*). To foreshadow, our model depicting relations among measures to predict these facets of situational trust is presented in Figure 2.

Two of our potential predictors can be thought of as baseline measures and appear in the leftmost column of Fig. 2 – overall Cyber Attack Knowledge and Pre-experimental Trust in Technology. These measures were positively related to each other (r = .242, p = .004) – i.e., more knowledgeable users had a higher baseline propensity to trust technology. In terms of predicting situational trust, greater knowledge unsurprisingly predicted lower situational trust (i.e., better recognition of threat; r = -.194, p = .021). However, *pre-experimental* trust was not a significant predictor of *situational trust* (r = -.058, p = .493, *ns*). This lack of relation was somewhat surprising because we expected that higher levels of baseline (pre-experimental) trust in technology might bias users to have higher situational levels of trust (in spite of threat cues).

We also expected that the affective responses (arousal and emotional valence) to the attack scenarios would predict situational trust. In all, as shown in the top right of Fig. 2, four factors predicted situational trust based on zero-order correlations: cyber knowledge, the overall arousal response (r = -.175, p = .037), the overall emotional valence response (r = .333, p < .001), and the risk sensitivity (|high – low|) of the emotional valance response (r = -.177, p < .035). Specifically, the lower (more negative) the emotional responses, and the greater the emotional risk sensitivity and the arousal, the lower the situational trust. A more detailed discussion of risk sensitivity measures will be presented in the next subsection.

Figure 2. *A Depiction of Relations Among Measures With an Ultimate Goal to Predict Situational Trust (Overall and Risk Sensitivity)*.



Note. r values are zero-order correlations (missing links reflect non-significant zero-order correlations). For the mediation analysis (blue triangle): TE=total effect, DE=direct effect, IE=indirect effect.

Meanwhile, to further develop a model to predict overall situational trust, we next sought to test a hypothesis that the impact of arousal on situational trust might be mediated by emotional valence. Physiological arousal (e.g., heart rate) sensations can be high for both positive and negative events, whereas an emotional valance response

also includes a cognitive attribution about the event (i.e., more negative valence suggests the event is bad/threatening). Thus, it seems plausible that this attribution (emotional valance) could mediate the impact of arousal on situational trust. The variables involved in this mediation analysis are connected via blue lines in Fig. 2, which also provides the r values of the Total Effect (TE), Direct Effect (DE) and Indirect Effect (IE) components. For this mediation analysis, we used the causal steps procedure (Baron & Kenny, 1986). The regression to check for the indirect effect of arousal on trust yielded an IE_a unstandardized coefficient of -.158 (*SE* = .063). The regression including both arousal and emotion to predict trust yielded an IE_b (indirect effect of emotion on trust) unstandardized coefficient of .278 (SE_b = 0.73). Significance of the mediation (i.e., that the effect of arousal on situational trust was mediated by emotional valance) was confirmed with a Sobel test = -2.09, *SE* = .021, *p* = .036. A regression analysis in which the emotional valance and arousal variables were entered sequentially also revealed that once the emotional valence predictor was included, the direct effect of arousal on situational trust was no longer significant (*t* = -1.364, *p* = .175), supporting an interpretation that emotional valance (completely) mediates the relation between arousal and situational trust. We do, however, acknowledge that there are limitations of mediation analyses – for example, they cannot strictly establish causation and there could be other key/latent variables missing from the model.

Our ultimate goal was to determine a regression model to predict situational trust. Besides knowledge and emotional valance (which mediates the effect of arousal), another zero-order predictor of situational trust was emotional risk sensitivity (|high – low|). However, this risk-sensitivity measure was also related to overall emotional valence (*r* = .277, *p* = .001), and a regression model to predict situational trust revealed that this factor became non-significant if entered after knowledge and overall emotional valence (*t* = 1.115, *p* = .267). Thus, in the model excluding non-significant factors, cyber attack knowledge predicted 4% of variance in overall situational trust, *F*(1,140) = 5.47, *p* = .021, and the emotional valence response predicted an additional 10% of variance beyond that explained by attack knowledge, *F*(2,139) = 10.70, *p* < .001. Knowledge was entered into the model first because it is a more long-term aspect of the participant and temporally existed before the temporary affective responses to the scenarios.

Risk Sensitivity

Overall responses collapsed across risk levels do not provide possibly relevant information about a user's risk sensitivity – that is, how much the user's response for a measure (e.g., arousal) varied as a function of the risk level manipulation in the scenarios (high vs. low). Thus, we operationalized risk sensitivity for each measure (arousal, emotion and situational trust) as the difference between the participant's responses for high-risk versus low-risk scenarios (i.e., high-low). Thus, we were also interested in predicting risk sensitivity in situational trust, and included emotional risk sensitivity in the model (Fig. 2). A finding not directly related to predicting a facet of situational trust was that both knowledge (*r* = .200, *p* = .017) and pre-experimental trust (*r* = .186, *p* = .027) predicted emotional risk sensitivity (whereas neither significantly predicted *overall* emotional valence). Thus, the more knowledgeable a participant, the more sensitivity they tended to exhibit to the risk manipulation in their emotional valence responses. This relationship with knowledge (or pre-experimental trust) was not significant for risk sensitivity in arousal or situational trust. However, risk sensitivity tended to be correlated across the affective measures - the larger an individual's risk sensitivity in emotional response, the larger was their risk sensitivity in arousal (*r* = .296, *p* < .001) and situational trust (*r* = .233, *p* = .005). However, risk sensitivity in arousal only marginally predicted risk sensitivity in situational trust (*r* = -.141, *p* = .094). Thus, as shown in Figure 2, emotional risk sensitivity was the sole significant predictor of risk sensitivity in situational trust. The correlations involving risk sensitivity measures (the grey boxes in Figure 2) were not ones we anticipated doing pre-experimentally. However, based on differences in emotional risk sensitivity between high and low knowledge participants in the supplementary ANOVAs (attack type × risk level × knowledge level), we wondered if knowledge and/or propensity to trust would predict risk sensitivity in emotion, arousal, and situational trust, and if risk sensitivity would be related among the latter measures. If a Bonferroni correction were used among this set of 9 comparisons, the relation in the model between emotional risk sensitivity and situational trust risk sensitivity remains significant, but the relations between Emotional risk sensitivity and knowledge and propensity to trust become non-significant.

Discussion

As in the Results section, we first discuss the impacts of our manipulations (risk level × attack type) and then we discuss the general relations among the affective measures and the predictors of pre-experimental trust and cyber attack knowledge. Lastly, we discuss possible limitations and future research.

Effects of Risk Level

The current study manipulated the severity dimension of risk while controlling for the likelihood dimension - i.e., the scenarios involved cyber attacks already in progress (likelihood = 1). Thus, we were able to isolate the affective effects of severity. Our findings can serve as a baseline for future work exploring the impact of variations in likelihood and/or dimensional interactions. Arousal, emotional valence and knowledge measures were all sensitive to the risk-level manipulation. While it may not be surprising that more severe scenarios produced increased arousal and more negative emotions than low-risk scenarios, these patterns provide some assurance that our risk manipulation in the form of scenarios was effective.

Interestingly, attack identification accuracy decreased as risk level increased. It is possible that the increased arousal (or more negative emotion) invoked by the high-risk scenarios might have impaired the cognitive attack identification process. Such impaired performance at high levels of arousal is compatible with the Yerkes-Dodson law (Cohen, 2011). Another possible cognitive reason (i.e., not mediated by affect) for poorer attack identification performance in high risk scenarios is that the participant's attention may have been divided between determining their best course of action to minimize loss in the scenario (though they were not explicitly required to state an action) and the task requirement to identify the type of attack in play, which may have been seen as a secondary low-priority goal (vs. as an *accuracy goal* in which to invest effort).

On the other hand, the emotional responses of high-knowledge users, who correctly identified all the attacks, exhibited more sensitivity to the risk manipulation than those of their less knowledgeable peers. Perhaps better understanding of the type of attack at hand allows more attention to be paid to the details of risk.

A somewhat unexpected result was that, in contrast to the other affective measures, there was no main effect of risk level on situational trust responses, which is reflected in the flat line for ransomware in Fig. 1c. However, an interaction with attack type revealed that situational trust did decrease with risk level as expected in the case of malware, but unexpectedly increased with risk level for phishing. In the high-risk phishing scenario, the email was purportedly from one's bank whereas in the low-risk scenario the email was purportedly from Apple regarding one's iTunes account. One possibility is that the mention of a bank automatically elicited a reaction of trust. Another possibility is that although all users likely had bank accounts (making an email from the bank seem somewhat plausible), not all users likely had iTunes accounts, which might make them more prone to consider an email from Apple as potentially suspect.

We also examined risk sensitivity as the difference in response across high- versus low- risk malware and ransomware scenarios. Risk sensitivity tended to be correlated across affective measures: thus, the more the increase in negative affect with risk, the more the increase in arousal, and the more the decrease in situational trust. It is possible that the decrease in situational trust was partly driven (mediated) by the increase in negative affect. Prior evidence suggests that negative affect can decrease trust in another individual (Dunn & Schweitzer, 2005), so it seems plausible that could also decrease situational trust in technology.

Effects of Attack Type

As hypothesized, phishing attacks lead to the least arousal and the least negative emotional valence (vs. ransomware or malware). As mentioned previously, a phishing attack affords more user control over the situation as it requires the user to voluntarily provide personal information, which can be avoided if the email or message is recognized as a phishing attempt. This explanation in terms of control is compatible with the finding that, in contrast, threats perceived as uncontrollable will elicit a high degree of dread (Slovic, 2016). This degree-of-control account is also compatible with the finding that greater cyber attack knowledge predicted a more positive emotional response for phishing (i.e., more knowledgeable users don't despair because they know they can thwart

the attack), whereas the trend was negative for malware and ransomware (i.e., the more knowledgeable the user, the more they tend to realize the damage is done and despair). This perceived personal locus of control may similarly explain the higher levels of *situational trust* in technology in response to phishing (relative to malware and ransomware). Our pattern of results is thus consistent with the hypothesis of Kostyuk and Wayne (2021) that knowledge and controllability influence assessments of cyber risk. There is also research to suggest that phishing attacks may instead/also negatively impact trust in other people ("phishers") or in oneself in the form of embarrassment and loss of self-confidence if one fell for the attack (Kelley et al., 2012).

Another (not exclusive) factor that may contribute to more muted emotional responses is familiarity. Note that phishing attacks may be familiar to most users due their frequent mention in the media and coverage in security awareness training campaigns (Proofpoint, 2020). If so, phishing attacks might have been the least shocking to participants because of familiarity with this type of attack. The general finding that familiarity can reduce one's affective response is the basis of desensitization therapy for phobias (Lazarus, 1961).

Besides familiarity gained through the media and/or training, many participants might also have had first-hand experience at being "phished" (e.g., received phishing emails). A study polling 155 participants on Mechanical Turk found that every single one of their participants had experienced phishing attempts (Kelley et al., 2012). We did not collect data on our participants' past experience with attacks, but if Kelley et al.'s findings generalize to our sample then all or nearly all our participants had experienced phishing attacks. One might expect that first-hand familiarity (in addition to familiarity due to media/training) might also contribute to a reduced affective response to phishing. However, research by Blum et al. (2014) suggests more generally that first-hand experience with negative events can actually increase (vs. decrease) one's current perception of risk (and thus, presumably, one's negative affective response). As an analogy, if a person had a first-hand experience of being attacked by a dog, one might expect that person to be more anxious around dogs than a person without that negative past experience. Thus, familiarity with attacks from training (a safe context) versus familiarity with attacks from first-hand experience might have different influences on emotional responses - with the former contributing to muted responses and the latter contributing to amplified responses. That said, a first-hand experience with phishing might not necessarily be especially negative because of the degree of user control associated with such attacks.

In all, phishing responses were somewhat distinctive (e.g., more muted affect) relative to malware and ransomware response patterns. This difference may be due to an increased personal locus of control and/or increased familiarity from training or first-hand experience if the experience was not emotionally negative (e.g., the user thwarted the phishing effort). Muted emotional responses are advantageous because over-arousal in response to a potential attack can be detrimental to effective decision making (Yerkes-Dodson Law: Cohen, 2011). To the extent that a contributor to the muted affective responses is an emphasis on phishing in training campaigns, training simulations that expose users to less common (but potentially more dangerous) types of attack scenarios may be able to take advantage of this familiarity effect to reduce subsequent arousal when users are exposed to those other types of attacks, and thus decrease affective impairment in response performance.

Although ransomware scenarios elicited stronger affective responses than phishing, the data did not provide support for our expectation that they would elicit stronger affective responses than the malwares scenarios. It is possible that because users knew that no actual money was on the line that the hypothetical scenarios did not evoke the sense of being robbed by another human being (vs. being inconvenienced by an impersonal computer virus). That said, ransomware scenarios did result in the least situational trust (Table 3). Further research is necessary to explore whether the salience of human agency, the analogy to physical robbery, and/or the addition of insult (ransom request) to injury (file inaccessibility) may heighten users' affective response to ransomware relative to other types of malware in the wild.

Individual Differences: Predicting Responses

More knowledgeable participants tended to have a higher general propensity to trust in technology. Thus, propensity to trust technology does not reflect being naive. Rather, less knowledgeable users may be inclined to generally distrust what they don't understand. In terms of situational trust, however, more knowledgeable participants reported lower trust levels in response to the cyber attack scenarios. Although correlation does not establish causation, a possible explanation for this pattern is that when people know enough to be aware they are

under attack, they likely adjust their vigilance and trust accordingly. Some participants were not even aware that all the scenarios constituted attacks as evidenced by responses like “none” or “nothing” to the attack-type knowledge question. In such cases, their situational trust levels were inappropriately high. Thus, more knowledgeable users seem able to appropriately adapt their trust level to the current context, which is a crucial skill given the dynamic demands of human-computer interaction in connected environments.

It was contrary to our expectations that unlike knowledge, pre-experimental trust levels did not predict situational trust levels. In retrospect it may make sense that just because one has a higher propensity to trust technology in general, one need not be assumed to be wearing rose colored glasses when faced with a current, specific attack. It was also surprising that increased knowledge was not associated with muted emotional or arousal responses. Indeed. emotional valence ratings were more sensitive to the manipulation of risk among more knowledgeable participants. Finally, also somewhat surprising was that correlations among the affective measures were not stronger. However, this suggests that they are not mutually redundant and that it may be important to include more than one affective measure in this type of research (i.e., emotional valence and arousal are distinct).

In terms of predicting situational trust, our model (Fig. 2) identified only cyber attack knowledge and emotional valence as unique predictors (as the effects of arousal and emotional risk sensitivity were fully mediated by emotional valence), and together these accounted for between 13% and 14% of the variance. Thus, additional factors are needed to further understand individual differences in situational trust.

Limitations and Future Research

Our predominantly male American college student sample could potentially limit the generalizability of our results. For example, older adults might tend to have less intense emotional responses to events than young adults (Birditt & Fingerman, 2003). Older age ranges (e.g., >40) may also tend to have lower levels of computer/cyber knowledge. That said, our sample was recruited from a pool including all possible majors at the institution, and yielded sufficient variability in factors like knowledge and trust in technology to reveal individual difference relations - e.g., greater knowledge was associated with increased propensity to trust technology. Future research is necessary to determine whether our findings extend to samples with different demographics (e.g., groups that are older and/or predominantly female).

In terms of quantifying arousal, a possible limitation of the current research is the use of self-report measures rather than possible physiological measures of arousal such as Galvanic Skin Response (GSR) (Rauch et al., 2014), Electrodermal Activity (EDA) (Braithwaite et al., 2013; Critchley, 2002), or cortisol levels (Backhaus et al., 2020; Canetti et al., 2017). That said, evidence suggests that the self-report emotional valence and arousal measures used in the current study do co-vary with physiological measures (Backhaus et al., 2020; Cuthbert et al., 2000; Lang et al., 1993).

Another possible limitation is that the cyber attack stimuli were operationalized as scenarios (rather than implementing them as unexpected pop-ups or cursor movements on participants’ computer screens while they performed a task). Nonetheless, as evident from Figure 1, the scenario manipulations successfully produced effects. With respect to the scenarios, we also acknowledge that the high versus low-risk pairs within each attack type involved somewhat different contexts and wordings, rather than having two almost identical scenarios for the high and low risk pairs with just a key word or two altered. One reason for this variation was that in our within-subjects design we felt it would be odd and redundant for a participant to see a stimulus set consisting of three pairs (albeit intermixed) of practically identical scenarios. That said, to reduce degrees of freedom, future work with between-subjects designs could use high/low scenarios with more closely corresponding wordings. Finally, all scenarios in the current study included an attack (though they varied in type and risk level). In practice, actual attacks may be a relative rarity in a stream of otherwise non-attack transactions, which could potentially reduce vigilance and hit rate in terms of attack recognition. The relatively good recognition rate in the current study may allow it to provide a baseline affective response profile in the case of largely successful attack recognition (see Table 3). Future studies may vary the proportion of attacks in the stimuli to explore the impact on recognition and affective response.

In our discussion of phishing, we suggested that familiarity might be a factor that influences affective responses. Thus, a possible extension to this research could involve explicitly adding measures to assess participants' familiarity (first-hand vs. via training) with different types of cyber attacks, because familiarity might mediate or moderate relations between other variables. As we did not survey participants as to whether they had had first-hand exposure to these types of attacks, we could not independently assess the impact of first-hand familiarity on affective responses (as distinct from the impact of knowledge and/or familiarity from training). Future work could also include other factors - such as measures of personality traits and expanded measures of cyber knowledge (beyond ability to identify attack type) - to determine their impact on users' affective responses and their relations with other factors (e.g., trust in automation).

Another possible avenue for future research would be to introduce a training manipulation. Prior research suggests that training on cyber attacks, specifically phishing and identifying fake websites, lowers susceptibility to cyber attacks (e.g., Sheng et al., 2007). It would be interesting to implement a training manipulation while simultaneously assessing changes in users' affective responses. The current correlation patterns suggest that more knowledgeable individuals tend to have a higher general trust in technology, but report lower situational trust when confronted with specific attacks. It would be interesting to assess if a training manipulation could establish this as a causal effect – that is, would training increase the general trust in technology, but also result in increased situational distrust in the face of an attack.

Another important avenue for further research would be to assess how exposure to personally relevant cyber threat scenarios and the resulting affective responses impact the subjects' future computer security attitudes and behaviors. When Kostyuk and Wayne's (2021) subjects were exposed to an article about a personally relevant cyber data breach, they reported a higher level of perceived personal risk for future cyber threats, but did not actually change their on-line behavior (i.e., remained susceptible to a phishing email) or their reported support for specific national cyber-security policies. In their study, however, subjects were not assessed directly on their emotional responses (valence & arousal) to their cyber threat stimuli, so it would be interesting to investigate if the nature or strength of affective responses might mediate/predict behavioral change and policy support.

Conclusions

Including multiple attack types in our design revealed that different attack types may yield different affective response patterns (i.e., phishing responses more muted relative to those for malware and ransomware). This result has the broader implication that findings from studies focused on a single type of attack, including the many studies on phishing, may not necessarily generalize to predict user responses to other types of cyber attacks.

The more knowledgeable the user, the lower their situational trust when faced with cyber attack scenarios, but the higher was their pre-experimental propensity to trust in technology. Thus, a general propensity to trust technology may not reflect naiveté but wisdom, and more knowledgeable users may be able to appropriately calibrate their trust level to the current context. In general, however, the affective responses to cyber attacks could not be thoroughly predicted by cyber attack knowledge and pre-experimental trust in technology (Fig. 2). Thus, further research must identify other factors to better predict and understand affective responses.

The different responses (arousal, emotional valence and situational trust) although correlated, were not so strongly correlated as to suggest they were mutually redundant. In the context of predicting situational trust, the emotional valence measure proved more relevant than the arousal measure (whose influence was fully mediated by emotional valence).

Footnotes

1. Concurrently with their university education, the students in our sample are also receiving training/instruction to enable them to serve as Army Officers upon graduation. In that capacity they can be referred to as cadets, so the use of the term cadet in the vignettes was intended to make them relatable to the participants.

2. Although ANOVAs have been considered robust to normality violations (Glass et al., 1972), since our Likert scale response measures were not always normally distributed, we also checked and confirmed all our main effects for

those measures using non-parametric Friedman Tests (for the 3-level attack type factor) or Wilcoxon Signed-rank tests (for the 2-level risk factor). Those results are provided as footnotes.

3. The main risk level effect was also checked/confirmed with a non-parametric Wilcoxon signed-ranks test, $Z = -4.55$, $p < .001$ (two-tailed).

4. The main attack type effect was also checked/confirmed with a non-parametric Friedman test, chi-squared (2, $N = 142$) = 9.33, $p = .009$.

5. The main risk level effect was also checked/confirmed with a non-parametric Wilcoxon signed-ranks test, $Z = -6.50$, $p < .001$ (two-tailed).

6. We confirmed this main effect of attack type using a non-parametric Friedman test, chi-square (2, $N = 142$) = 41.26, $p < .001$.

7. The absolute value |high-low| was used to make the results easier to interpret. Otherwise for just (high-low), more negative (lower) numbers would reflect greater risk sensitivity for emotional valance and trust.

References

Agrafiotis, I., Nurse, J. R. C., Goldsmith, M., Creese, S., & Upton, D. (2018). A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate. *Journal of Cybersecurity, 4*(1), Article tyy006. https://doi.org/10.1093/cybsec/tyy006

Aycock, J., Crawford, H., & deGraaf, R. (2008). Spamulator: The Internet on a laptop. *ACM SIGCSE Bulletin*, *40*(3), 142–147. https://doi.org/10.1145/1597849.1384311

Backhaus, S., Gross, M. L., Waismel-Manor, I., Cohen, H., & Canetti, D. (2020). A cyberterrorism effect? Emotional reactions to lethal attacks on critical infrastructure. *Cyberpsychology, Behavior, and Social Networking*, *23*(9), 595–603. https://doi.org/10.1089/cyber.2019.0692

Baker, M. (2016). *Striving for effective cyber workforce development.* Software Engineering Institute.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Birditt, K. S., & Fingerman, K. L. (2003). Age and gender differences in adults' descriptions of emotional reactions to interpersonal problems. *The Journals of Gerontology: Series B*, *58*(4), P237–P245. https://doi.org/10.1093/geronb/58.4.P237

Blum, S. C., Silver, R. C., & Poulin, M. J. (2014). Perceiving risk in a dangerous world: Associations between life experiences and risk perceptions. *Social Cognition*, *32*(3), 297–314. https://doi.org/10.1521/soco.2014.32.3.297

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Behavioral Therapy and Experimental Psychiatry, 25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). *A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments*. University of Birmingham.

Buck, R., Khan, M., Fagan, M., & Coman, E. (2017). The User Affective Experience Scale: A Measure of emotions anticipated in response to pop-up computer warnings. *International Journal of Human–Computer Interaction*, *34*(1), 25–34. https://doi.org/10.1080/10447318.2017.1314612

Canetti, D., Gross, M., Waismel-Manor, I., Levanon, A., & Cohen, H. (2017). How cyberattacks terrorize: Cortisol and personal insecurity jump in the wake of cyberattacks. *Cyberpsychology, Behavior, and Social Networking*, *20*(2), 72–77. https://doi.org/10.1089/cyber.2016.0338

Chen, J. Y. C., & Terrence, P. I. (2009). Effects of imperfect automation and individual differences concurrent performance of military robotics tasks in a simulated multitasking environment. *Ergonomics*, *52*(8), 907–920. https://doi.org/10.1080/00140130802680773

Chen, Y., Zahedi, F. M., Abbasi, A., & Dobolyi, D. (2021). Trust calibration of automated security IT artifacts: A multi-domain study of phishing-website detection tools. *Information & Management*, *58*(1), Article 103394. https://doi.org/10.1016/j.im.2020.103394

Chong, I., Xiong, A., & Proctor, R. W. (2019). Human factors in the privacy and security of the internet of things. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *27*(3), 5–10. https://doi.org/10.1177/1064804617750321

Chughtai, A. A., & Buckley, F. (2008). Work engagement and its relationship with state and trait trust: A conceptual analysis. *Journal of Behavioral and Applied Management*, *10*(1), 47–71. https://doi.org/10.21818/001c.17170

Cohen, R. A. (2011). Yerkes–Dodson Law. In *Encyclopedia of clinical neuropsychology* (pp. 2737–2738). Springer.

Pricewaterhouse Coopers. (2016). *Turnaround and transformation in cybersecurity: Key findings from The Global State of Information Security Survey 2016*. https://www.pwc.com/sg/en/publications/assets/pwc-global-state-of-information-security-survey-2016.pdf

Critchley, H. D. (2002). Electrodermal responses: What happens in the brain. *The Neuroscientist*, *8*(2), 132–142. https://doi.org/10.1177/107385840200800209

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective picture processing: Covariation with autonomic arousal and affective report. *Biological Psychology*, *52*(2), 95–111. https://doi.org/10.1016/S0301-0511(99)00044-7

de Visser, E., Shaw, T., Mohamed-Ameen, A., & Parasuraman, R. (2010). Modeling human-automation team performance in networked systems: Individual differences in working memory count. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*(14), 1087–1091. https://doi.org/10.1177/154193121005401408

Dhamija, R., Tygar, J. D., & Hearst, M. (2006). Why phishing works. In *CHI '06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 581–590). ACM. https://doi.org/10.1145/1124772.1124861

Downs, J. S., Holbrook, M., & Cranor, L. F. (2007). Behavioral response to phishing risk. In *eCrime '07: Proceedings of the Anti-phishing Working Groups 2nd Annual eCrime Researchers Summit* (pp. 37–44). ACM. https://doi.org/10.1145/1299015.1299019

Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and believing: The influence of emotion on trust. *Journal of Personality and Social Psychology*, *88*(5), 736–748. https://doi.org/10.1037/0022-3514.88.5.736

Dutt, V., Ahn, Y.-S., & Gonzalez, C. (2013). Cyber situation awareness: Modeling detection of cyber attacks with instance-based learning theory. *Human Factors*, *55*(3), 605–618. https://doi.org/10.1177/0018720812464045

Endsley, M. R., & Kiris E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, *37*(2), 381–394. https://doi.org/10.1518/001872095779064555

Fagan, M., Khan, M. M. H., & Buck, R. (2015). A study of users' experiences and beliefs about software update messages. *Computers in Human Behavior*, *51*(Part A), 504–519. https://doi.org/10.1016/j.chb.2015.04.075

Faul, F., & Erdfelder, E. (1992). *GPOWER: A priori, post-hoc, and compromise power analyses for MS-DOS* [Computer program]. Bonn University.

Gilles, I., Bangerter, A., Clémence, A., Green, E. G. T., Krings, F., Staerklé, C., & Wagner-Egger, P. (2011). Trust in medical organizations predicts pandemic (H1N1) 2009 vaccination behavior and perceived efficacy of protection measures in the Swiss public. *European Journal of Epidemiology*, *26*(3), 203–210. https://doi.org/10.1007/s10654-011-9577-2

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, *42*(3), 237–288. https://doi.org/10.3102/00346543042003237

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Gomez, M. A. (2019). Past behavior and future judgements: Seizing and freezing in response to cyber operations. *Journal of Cybersecurity, 5*(1), Article tyz012. https://doi.org/10.1093/cybsec/tyz012

Gross, M. L., Canetti, D., & Vashdi, D. R. (2016). The psychological effects of cyber terrorism. *Bulletin of the Atomic Scientists, 72*(5), 284–291. https://doi.org/10.1080/00963402.2016.1216502

Gross, M. L., Canetti, D., & Vashdi, D. R. (2017). Cyberterrorism: Its effects on psychological well-being, public confidence and political attitudes. *Journal of Cybersecurity*, *3*(1), 49–58. https://doi.org/10.1093/cybsec/tyw018

Hess, A. (2015, November 22). "Everything was completely destroyed": What it was like to work at Sony after the hack. *Slate*. http://www.slate.com/articles/technology/users/2015/11/sony_employees_on_the_hack_one_year_later.html

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Kelley, C. M., Hong, K. W., Mayhorn, C. B., & Murphy-Hill, E. (2012). Something smells phishy: Exploring definitions, consequences, and reactions to phishing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 2108–2112. https://doi.org/10.1177/1071181312561447

Kincaid, J. P., Fishburne, R. P., Jr., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel* (Report no. RBR-8-75). Institute for Simulation and Training, University of Central Florida. https://stars.library.ucf.edu/istlibrary/56/

Koelsch, S., Kilches, S., Steinbeis, N., & Schelinski, S. (2008). Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity. *Plos One*, *3*(7), Article e2631. https://doi.org/10.1371/journal.pone.0002631

Kostyuk, N., & Wayne, C. (2021). The microfoundations of state cybersecurity: Cyber risk perceptions and the mass public. *Journal of Global Security Studies, 6*(2), Article ogz077. https://doi.org/10.1093/jogss/ogz077

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*(3), 261–273. https://doi.org/10.1111/j.1469-8986.1993.tb03352.x

Lazarus, A. A. (1961). Group therapy of phobic disorders by systematic desensitization. *The Journal of Abnormal and Social Psychology*, *63*(3), 504–510. https://doi.org/10.1037/h0043315

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, *40*(1), 153–184. https://doi.org/10.1006/ijhc.1994.1007

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50.30392

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.2307/258792

Merritt, S. M., Heimbaugh, H., LaChapell, J., & Lee, D. (2013). I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. *Human Factors*, *55*(3), 520–534. https://doi.org/10.1177/0018720812465081

Mosier, K., Fischer, U., & The HART Group. (2012). Impact of automation, task and context features on pilots' perception of human-automation interaction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 70–74. https://doi.org/10.1177/1071181312561035

Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation bias: Decision making and performance in high-tech cockpits. *The International Journal of Aviation Psychology, 8*(1), 47–63. https://doi.org/10.1207/s15327108ijap0801_3

Mooradian, T., Renzl, B., & Matzler, K. (2006). Who trusts? Personality, trust and knowledge sharing. *Management Learning*, *37*(4), 523–540. https://doi.org/10.1177/1350507606073424

Nieto, A., Rios, R. (2019). Cybersecurity profiles based on human-centric IoT devices*. Human-centric Computing and Information Sciences*, *9*, Article 39. https://doi.org/10.1186/s13673-019-0200-y

Pak, R., McLaughlin, A. C., & Bass, B. (2014). A multi-level analysis of the effects of age and gender stereotypes on trust in anthropomorphic technology by younger and older adults. *Ergonomics, 57*(9), 1277–1289. https://doi.org/10.1080/00140139.2014.928750

Panda Labs. (2016, October 16). *Cybercrime reaches new heights in the third quarter*. https://www.pandasecurity.com/mediacenter/pandalabs/pandalabs-q3/

Parasuraman, R., de Visser, E., Lin, M.-K., & Greenwood, P. M. (2012). Dopamine beta hydroxylase genotype identifies individuals less susceptible to bias in computer-assisted decision making. *Plos One*, *7*(6), Article e39675. https://doi.org/10.1371/journal.pone.0039675

Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. *Human Factors, 50*(3), 511–520. https://doi.org/10.1518/001872008X312198

Pfleeger, S. L., & Caputo, D. D. (2012). Leveraging behavioral science to mitigate cyber security risk. *Computers & Security, 31*(4), 597–611. https://doi.org/10.1016/j.cose.2011.12.010

Prati, G., Pietrantoni, L., & Zani, B. (2011). Compliance with recommendations for pandemic influenza H1N1 2009: The role of trust and personal beliefs. *Health Education Research*, *26*(5), 761–769. https://doi.org/10.1093/her/cyr035

Proofpoint. (2020). *State of the phish: An in-depth look at user awareness, vulnerability and resilience* (Annual Report). https://www.proofpoint.com/us/resources/threat-reports/state-of-phish

Rauch, S. M., Strobel, C., Bella, M., Odachowski, Z., & Bloom, C. (2014). Face to face versus Facebook: Does exposure to social networking web sites augment or attenuate physiological arousal among the socially anxious? *Cyberpsychology, Behavior, and Social Networking*, *17*(3), 187–190. https://doi.org/10.1089/cyber.2012.0498

Rossi, P. H., & Anderson, A. B. (1982). The factorial survey approach: An introduction. In P. H. Rossi & S. L. Nock (Eds.), *Measuring social judgments: The factorial survey approach* (pp. 15–67). Sage.

Rovira, E., McLaughlin, A. C., Pak, R., & High, L. (2019). Looking for age differences in self-driving vehicles: Examining the effects of automation reliability, driving risk, and physical impairment on trust. *Frontiers in Psychology*, *10*, Article 800. https://doi.org/10.3389/fpsyg.2019.00800

Rovira, E., Pak, R. & McLaughlin, A. (2017). Effects of individual differences in working memory on performance and trust with various degrees of automation. *Theoretical Issues in Ergonomics Science*, *18*(6), 573–591. https://doi.org/10.1080/1463922X.2016.1252806

Sawyer, B. D., & Hancock, P. A. (2018). Hacking the human: The prevalence paradox in cybersecurity. *Human Factors, 60*(5), 597–609. https://doi.org/10.1177/0018720818780472

Sheng, S., Chan, W. L., Li, K. K., Xianzhong, D., & Xiangjun, Z. (2007). Context information-based cyber security defense of protection system. *IEEE Transactions on Power Delivery*, *22*(3), 1477–1481. https://doi.org/10.1109/TPWRD.2006.886775

Slovic, P. (2016). *The perception of risk*. Routledge.

SonicWall (2021). *Mid-Year Update Cyber Threat Report*. https://www.sonicwall.com/medialibrary/en/white-paper/mid-year-2021-cyber-threat-report.pdf

Wogalter, M. S., Young, S. L., Brelsford, J. W., & Barlow, T. (1999). The relative contributions of injury severity and likelihood information on hazard-risk judgments and warning compliance. *Journal of Safety Research*, *30*(3), 151–162. https://doi.org/10.1016/S0022-4375(99)00010-9

Yu, K., Taib, R., Butavicius, M. A., Parsons, K., & Chen, F. (2019). Mouse behavior as an index of phishing awareness. In D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, & P. Zaphiris (Eds.), *Human-Computer Interaction – INTERACT 2019*. Springer.

Correspondence to:
Aryn Pyke
Associate Professor & Research Scientist
Army Cyber Institute, West Point
2101 New South Post Road
Highland Falls, NY 10928
United States
Email: aryn.pyke(at)westpoint.edu

Editorial record: First submission received on May 20, 2020. Revisions received on March 10, 2021, October 1, 2021 and October 13, 2021. Accepted for publication on October 14, 2021.
Editor in charge: David Smahel

About Authors

Aryn Pyke, Ph.D., is an Associate Professor in the Engineering Psychology Program at West Point, and a Cognitive Scientist with the Army Cyber Institute. She holds bachelors and master's degrees in electrical and computer engineering and a PhD in Cognitive Science. Her research interests include STEM education and human-computer interaction. Her current focus is on the human-in-the-loop in cybersecurity and AI-human teaming contexts.

Ericka Rovira, Ph.D., is a Professor of Engineering Psychology in the Department of Behavioral Sciences and Leadership at the United States Military Academy, West Point, NY. Her research focuses on human autonomy teaming in high-risk complex environments. Her expertise lies in understanding how to improve trust and reliance in robots and other autonomous systems. Dr. Rovira is a Fellow and Past President of the American Psychological Association's Division 21: Applied Experimental and Engineering Psychology Division.

CPT **Savannah Murray** is a US Army Logistics Officer and a 2017 graduate from the U.S. Military Academy. She graduated from the U.S. Army Ranger school in 2019. She helped conduct this research as part of her undergraduate senior thesis.

CPT **Joseph Pritts** is a U.S. Army Engineer Officer and a 2017 graduate from the U.S. Military Academy. He helped conduct this research as part of his undergraduate senior thesis.

Charlotte L. Carp, Ph.D., is an Assistant Clinical Professor at the University of Huston in the Department of Educational Leadership & Policy Studies. Her current research focus is on language development and special populations, with the objective of making language interventions more effective and efficient for persons with disabilities.

Robert Thomson, Ph.D., serves as the Cyber and Cognitive Science Fellow at the Army Cyber Institute and is an Associate Professor of Engineering Psychology in the Department of Behavioral Sciences and Leadership at West Point. Dr. Thomson's current research interests are at the intersection of computational modeling, cybersecurity, and artificial intelligence.