CONTRIBUTED PAPER

Internet-Based Social Engineering Psychology, Attacks, and Defenses: A Survey

This article systemizes Internet-based social engineering attacks through a psychological lens and investigates why current defenses have limited success. It also provides a roadmap for future research studies.

By Theodore Tangie Longtchi, Rosana Montañez Rodriguez, Laith Al-Shawaf, Adham Atyabi, *Member IEEE*, and Shouhuai Xu, *Senior Member IEEE*

ABSTRACT | Internet-based social engineering (SE) attacks are a major cyber threat. These attacks often serve as the first step in a sophisticated sequence of attacks that target, among other things, victims' credentials and can cause financial losses. The problem has received mounting attention in recent years, with many publications proposing defenses against SE attacks. Despite this, the situation has not improved. In this article, we aim to understand and explain this phenomenon by investigating the root cause of the problem. To this end, we examine Internet-based SE attacks and defenses through a unique lens based on psychological factors (PFs) and psychological techniques (PTs). We find that there is a key discrepancy between attacks and defenses: SE attacks have deliberately exploited 46 PFs and 16 PTs in total, but existing defenses have only leveraged 16 PFs and seven PTs in total. This discrepancy may explain why existing defenses have achieved limited success and prompt us to propose a systematic roadmap for future research.

KEYWORDS | Cyberattacks; deception; email-based attacks; Internet attacks; online social network (OSN)-based attacks; phishing; psychological factors (PFs); psychological techniques (PTs); social engineering (SE) attacks; website-based attacks.

Manuscript received 20 June 2023; revised 2 January 2024; accepted 14 March 2024. Date of publication 5 April 2024; date of current version 1 May 2024. This work was supported in part by NSA under Grant 43000871, in part by NSF under Grant 2115134 and Grant 2308142, and in part by Colorado State Bill 18-086. *(Corresponding author: Shouhuai Xu.)*

Theodore Tangie Longtchi, **Adham Atyabi**, and **Shouhuai Xu** are with the Department of Computer Science, University of Colorado Colorado Springs, Colorado Springs, CO 80918 USA (e-mail: sxu@uccs.edu).

Rosana Montañez Rodriguez is with the Department of Computer Science, The University of Texas at San Antonio, San Antonio, TX 78249 USA.

Laith Al-Shawaf is with the Department of Psychology, University of Colorado Colorado Springs, Colorado Springs, CO 80918 USA, and also with the Institute for Advanced Study in Toulouse (IAST), 31080 Toulouse, France.

Digital Object Identifier 10.1109/JPROC.2024.3379855

I. INTRODUCTION

Humans are the weakest link in cybersecurity, and this situation is seemingly worsening. This can be evidenced by an FBI report stating a $26B loss between June 2017 and July 2019 associated with attack emails that contain instructions on approving payments to attackers while pretending to come from executives [1] and another FBI report [2] stating that the financial loss increased to $43B from 2019 to 2022 during the COVID-19 pandemic, perhaps partly because most employees were working remotely and communications were mostly electronic rather than physical. Consider the example of phishing attacks, which are perhaps the most proliferated [3], the most investigated [4], and the most successful type of attack in causing security breaches [5]. The Anti-Phishing Working Group (APWG), which is arguably the organization that collects the most phishing emails in the world, reports that phishing has continued to grow since 2019; there were 0.8 million phishing websites in 2019, 1.8 million in 2020, 2.8 million in 2021, and more than 4.7 million in 2022 [3].

The increasingly significant damage caused by Internet-based social engineering (SE) attacks suggests that the tremendous efforts invested into designing

defenses against them appear to have achieved very limited success. To understand why this is the case, we aim to take a deeper look into the following problems: 1) what is the root cause that enables Internet-based SE attacks? 2) why have existing defenses achieved very limited success in mitigating these attacks? and 3) what research needs to be conducted in order to adequately mitigate these attacks? To answer these questions, we focus on three major classes of SE attacks: email-based, website-based, and online social network (OSN)-based SE attacks.

Our Contributions: We systematize Internet-based SE attacks and defenses through a psychological lens centered on the notion of psychological factors (PFs), which refer to the human attributes that can be exploited by attackers (i.e., *what to exploit*) and the notion of psychological techniques (PTs), which refer to the strategies that can be used to exploit PFs to encourage individuals to comply with an Internet-based SE attack (i.e., *how to exploit*). Specifically, we make the following five contributions.

First, we systematize the human PFs that have been exploited by attackers to wage attacks. We consider both the PFs that are *explicitly* discussed (i.e., elaborated) in the literature and the PFs that are *implicitly* discussed in the literature (i.e., mentioned but not elaborated). Similarly, we systematize the PTs that are explicitly or implicitly discussed in the literature. As highlighted in Table 1, we systematize 46 PFs (including 11 initial PFs from well-established psychological principles) and 16 PTs (including 3 initial PTs from literature); the 46 PFs are divided into five classes: *social psychological* PFs, *personality and individual difference* (PID) PFs, *cognitive* PFs, *emotion* PFs, and *workplace* PFs. These represent the most comprehensive list of PFs and PTs for future study while noting that humans are susceptible to SE attacks because of the PFs.

Second, we systematize Internet-based SE attacks while emphasizing the PFs and PTs that they exploit. We categorize attacks based on their *objectives* and *types*. As highlighted in Table 1, we cover nine attack objectives and 26 attack types, which represent the most comprehensive list of attack objectives and types that have been described in the literature. For each attack type, we summarize the PFs and the PTs that it leverages to exploit the PFs. Among the three classes of attacks, we report that email-based attacks have exploited 44 PFs through 11 PTs, website-based attacks have exploited 38 PFs through 12 PTs, and OSN-based attacks have exploited 41 PFs through 12 PTs. These suggest that attackers have been very aggressive in identifying and exploiting PFs using PTs.

Third, we systematize the defenses that take PFs or PTs into consideration, typically via feature definitions when applying machine learning techniques. As shown in Table 1, the number of defenses considering PFs or PTs is very small; there are only 12 defenses that consider PFs or PTs. We highlight that the state-of-the-art defenses have not adequately leveraged PFs or PTs because they collectively only consider 16 PFs and seven PTs, which are in sharp contrast to the 46 PFs and 16 PTs that have been exploited by SE attacks. This discrepancy may explain why current defenses have achieved limited success.

Fourth, we systematize the relationships between PFs, PTs, attacks, and defenses by mapping them. Throughout this article, we point to many findings, such as: 1) humans are inherently vulnerable to SE attacks, and the AUTHORITY PF is most exploited by SE attacks, followed by the TRUST, NEGLIGENCE, COGNITIVE MISER, FEAR, and GREED PF; 2) the attention grabbing PT is perhaps most exploited by SE attacks, but the persuasion PT is most studied; 3) money is the most popular attack objective; 4) business email compromise (BEC) attacks cause the largest financial losses, but the impersonation PT is most exploited by website-based SE attacks; 5) the personalization PT has exploited most PID PFs; and 6) PFs and PTs have not been adequately leveraged to design defenses, but it may be difficult to leverage the affection trust and quid-pro-quo PTs and the TRUST, IMPULSIVITY, CURIOSITY, FEAR, and NEGLIGENCE PFs when designing defenses.

Fifth, we propose a roadmap to guide future studies. The roadmap includes a systematic framework that describes the conceptual relationships between the relevant psychological concepts including the PFs and PTs that affect the effectiveness of defenses against Internet-based SE attacks. In particular, the framework can accommodate multiple psychological lenses (while noting that this study is centered at the lens of PFs and PTs) and seeks to quantitatively characterize the roles played by the PFs and PTs. The framework leads to specific approaches to designing future defenses, including: 1) design training schemes that accommodate the PFs and PTs exploited by attackers and 2) design automated defenses to adequately accommodate PFs and PTs.

Related Work: We focus on Internet-based SE attacks, in contrast to their counterpart in the physical world [49]. Moreover, we take a unique psychological lens, involving *psychology* in terms of PFs and PTs, *attacks* exploiting PFs or PTs, and *defenses* leveraging PFs or PTs. Note that there are many studies on defenses against Internet-based SE attacks, but most of them do not consider PFs or PTs (see [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], and [40]).

Table 1 highlights the comparison between existing surveys and ours. To further help understand the comparison, we elaborate on some examples. Khonji et al. [40] survey phishing definitions and detection methods; by contrast, we look at PFs and PTs. In terms of defenses, Guo et al. [27] present a comprehensive survey on Internet-based defenses, and Chanti and Chithralekha [23] present a comprehensive survey on defenses against phishing attacks, but they both only consider few PFs. Other surveys (e.g., [12], [13], [23], and [30]) consider more PFs but are less comprehensive than ours. Moreover, our

Table 1 Comparison Between Existing Surveys and Ours, Where the References Are Listed in the Chronicle Order, Defenses Are Divided Into Three Classes Based on the Attacks That They Are Defending Against (i.e., Email-Based Versus Website-Based Versus OSN-Based), an Empty Cell Means That the Reference Does Not Address the Issue in Question, a Number in Parentheses Means the Number of PFs/PTs/Attack Objectives/Attack Types That Are *Implicitly* Discussed (i.e., Not Elaborated) in a Reference, And a Number Without Parenthesis Means the Number of PFs/PTs/Attack Objectives/Attack Types That Are *Explicitly* Discussed (i.e., Elaborated) in the Reference. For Defenses, We Only Consider the Ones That Explicitly Considered/Leveraged PFs and/or PTs. We Observe That Several Prior Surveys Only Focus on One Attack Type. We Also Observe That Defenses Leveraging PFs/PTs Are Rare, as We Only Identify Four Defenses (Leveraging Ten PFs in Total as Indicated by the Number in the Brackets) Against Email-Based Attacks, Five Defenses (Leveraging Nine PFs in Total) Against Website-Based Attacks, and Three Defenses (Leveraging Nine PFs in Total) Against OSN-Based Attacks. Note That Only Our Article Systematizes the Relationships Between Attacks, Defenses, PFs, and PTs

Reference	Psychology		Attacks		Defenses		
Paper title and Reference	PFs	PTs	Objectives	Type	Email	Website	OSN
Review of intelligent detection designs of HTML URL phishing attacks [6]		(1)	(3)	1			
Review of phishing websites detection approaches [7]	(1)	(2)	(2)	1			
Impact of (in)formal organizational norms on susceptibility to phishing [8]	(3)	(2)	(4)	1			
Deep learning for phishing detection systematic literature review [4]		(1)	(2)	1			
Social engineering attacks prevention systematic literature review [9]	(15)	(4)	(3)	8		2	1
Impact of social engineering attacks: A literature review [10]	3		(2)	2(9)	1		
A study on the psychology of SE-based cyberattacks & countermeasures [11]	2(9)	6	(4)	13	3	1	1
Review of Social media identity deception detection [12]	(3)	(2)		7			2
Phishing techniques, defence mechanisms and open research challenges [13]	(5)	(4)	3	8	2	2	1
Machine-learning based Phishing detection review [14]	(2)	(1)	(2)	1			
A literature review of social engineering based on COVID-19 pandemic [15]		(1)	3(5)	3(8)	1		
Social engineering attacks: Recent advances and challenges [16]	(2)	2(1)	(3)	7(2)	3	1	1
Taxonomy of website anti-phishing solutions [17]	(1)	(1)	(3)	5			
Deceptive phishing attacks in social networking environments scrutiny [18]			(2)	1			
Heuristic-based strategy for phishing prediction using URL-based approach [19]		(2)		3			
Phishing attacks Types, vectors, and technical approaches review [20]	7 (12)	(3)	(3)	18		2	
Web phishing detection techniques taxonomy and future directions [21]	(10)	(2)	(3)	2		2	
Comprehensive review of effective anti-phishing training [22]	(18)	(4)	(2)	1	1		
Comprehensive classification of anti-phishing solutions [23]		(3)	(2)	1	2	2	
E-mail-based phishing attack taxonomy [24]	(2)	(3)	4	(5)			
Taxonomy of social engineering defense mechanism [25]	(6)	(3)	(2)	6	1	1	
AI-enabled phishing attacks detection techniques comprehensive review [26]		(2)	(1)	1		2	
Review of Online social deception and its countermeasures [27]	(3)	(3)	7	5			
Comprehensive reexamination of phishing research security perspective [28]	(6)	(3)	(5)	2 (5)			
Review of Social engineering studies and a study of attack scenarios [29]	18	4 (3)	(2)	11		1	
Social engineering attacks classification, detection and prevention [5]	(3)	3	(2)	14			
Classification of online social networks attacks and defence mechanism [30]	(3)	(1)	(2)	15			
Taxonomy of phishing attacks defense methods, issues, future directions [31]		(2)	6	4	1	2	
Phishing attacks types, vectors and technical approaches review [32]	(14)	(3)	(3)	11			
Review for Systematically understanding the cyber attack business [33]	(3)		2 (3)	(3)			
The impact of personality traits on user's susceptibility to SE attacks [34]	5(9)	(2)		1			
Systematic review of software-based web phishing detection [35]		(1)	(3)	2			
Phishing environments, techniques, and countermeasures review [36]			(5)	2 (4)	1		
Literature review of social engineering attacks with focus on Phishing [37]			(1)	1		1	
Semantics for social engineering attacks and defence mechanisms [38]	(2)	(1)	(2)	20			
Review of Phishing attacks and classification of emails [39]		(2)	(3)	4			
Literature review of Phishing detection [40]		(1)	3	(3)		2	
Phishing email filtering techniques review [41]			(2)	(1)			
This paper	46	16	9	26	4 [10]	5 [9]	3 [9]

study leads to new aspects that are not known until now, including: 1) mapping from SE attacks to PFs through the "bridge" of PTs; 2) defenders largely lagging behind attackers in leveraging PFs, explaining the limited success of current defenses; and 3) a systematic framework to guide the design and development of effective defenses. To elaborate item 3), we use Table 2 to highlight the comparison between the existing frameworks that consider psychological principles [42], [43], [44], [45], [46], [47], [48], [50], and ours. For example, [42], which is inspired by this study, aims to quantify the sophistication of emails based on the PFs and PTs that they exploit. This manifests the potential of the present survey to inspire future studies. As another example, [44] presents a human cognition framework to accommodate SE attacks while considering 22 PFs; in contrast, we consider 46 PFs and 16 PTs. As yet another example, [43] also presents a cognitive framework to dissect and characterize SE attacks but considers only four PFs.

In order to see the novelty in our framework described in our roadmap for future research, Table 2 compares seven existing frameworks and ours. The main difference between our framework and the others is the degree of *comprehensiveness* in terms of covering PFs, PTs, attacks, and defenses. For example, our framework is the only one that simultaneously accommodates the following aspects: *human information processing* (heuristic versus analytic), *risk attitude*, *individual baseline* (including the 46 PFs and 16 PTs identified in this study), *attack effort*, and *defense alerts*. Moreover, we stress the importance of establishing quantitative characteristics.

Last but not least, we are made aware of one very recent study [50], which investigates the design of empirical SE studies, especially the coverage of cognitive factors in

Table 2 Comparison Between Existing Frameworks and Ours Considering Psychological Principles

Reference	Concept/description	Psychology	Attacks	Defense	Survey?
Quantifying psychological sophistication of malicious emails [42]	A framework for quantifying sophistication of malicious emails by deconstructing their low-level and high-level psychological features	Leveraging psychological techniques (e.g., low-level psychological textual/imagery elements in emails) and psychological tactics (e.g., high-level assessment of attacker's effort as exhibited by email content)	Email-based SE attacks (e.g., Phishing, Scams and Spams)	Quantified sophistication can be leveraged to guide the design of effective defenses	No
Dissecting SE attacks through the lenses of cognition [43]	A framework for dissecting and characterizing SE attacks via cognitive features	Leveraging cognitive psychology concepts (e.g., Working Memory, Dual-processing models, Expert Utility) to understand the intruder-persuasion-dupe	SE attacks (e.g., Spear Phishing)	Guide the formulation and design of policies and training against SE attacks	No
Human cognition through the lens of SE attacks [44]	A framework for understanding SE attacks through the lens of human cognition and persuasion	Leveraging human factors such as perception, working memory, decision making, and action, workload, personality, expertise, and culture	SE attacks (e.g., Phishing, Water Holing, and Scams)	Guide the design of psychologically-principled defenses against SE attacks	No
The SE personality framework [45]	A framework for linking personality traits to principles of persuasion	Leveraging the Big Five Personality Traits and the Principle of Persuasion	SE attacks in the physical domain	Guide the design of awareness training schemes	No
Dissecting SE [46]	A framework for understanding SE attacks via the intruder-persuasion-dupe	Leveraging persuasion, fabrication (i.e., impersonation or providing misleading cues), and data gathering (e.g., dumpster diving, phishing)	SE attacks (e.g., Phishing, shoulder surfing)	Guide the design of of defenses against SE attacks	No
MINDSPACE: influencing behaviour for public policy [47]	Extending the 4E (Enable, Encourage, Engage and Exemplify) policy framework to 6E (adding Explore and Evaluate) to help policy-making	Leveraging MINDSPACE behaviors (i.e., Messenger, Incentives, Norms, Defaults, Salience, Priming, Affect, Commitments, and Ego)	Crimes to the public	Guide the design of polices to fight crimes	No
A framework for reasoning humans in the loop [48]	A framework for understanding the behavior of humans in performing security-critical functions	Leveraging the Communication-Human Information Processing (C-HIP) model with respect to five tasks (i.e., warnings, notices, status indicators, training, and policies)	Human vulnerabilities in security systems	Remove humans from the loop when designing security-critical functions	No
This paper	A framework for understanding Internet-based SE attacks through a psychological lens	Leveraging the Big Five Personality traits, the Principles of Persuasion as a starting point in identifying PFs and PTs	Systematizing 26 types of SE attacks	Envisioning a roadmap of future research directions	Yes

their experimental designs. By contrast, we systematically identify the PFs and PTs that can be used by SE attacks while going beyond cognitive factors because these factors only represent one class of PFs (out of the five classes that we consider).

Article Outline: Section II describes our methodology. Section III defines the psychological lens, whereby we conduct the study. Section IV shows how we identify the relevant literature. Section V systematizes PFs. Section VI systematizes PTs. Section VII systematizes attacks. Section VIII systematizes defenses. Section IX systematizes the relationships between PFs, PTs, attacks, and defenses. Section X presents a roadmap for future research. Section XI concludes this article.

II. GENERAL METHODOLOGY AND INSTANTIATION

Terminology: In this article, the term "social engineering (SE) attacks" or simply "attacks" refers to "Internet-based SE attacks" unless explicitly stated otherwise. The terms "individuals," "users," and "humans" are used interchangeably. The term "victims" refers to the users that are compromised by SE attacks. The term "SE defenses" or simply "defenses" refers to defenses against SE attacks.

General Methodology: Fig. 1 highlights the general methodology, which can accommodate any psychological lens of interest. This is important because different lenses would lead to findings that can be incorporated into a holistic understanding of SE attacks and defenses. The methodology is geared toward addressing eight research questions. More specifically, the methodology can be understood as follows.

First, the methodology is centered on the notion of *psychological lens*, which defines a unique psychological perspective through which we can systematize SE attacks and defenses. We propose defining a psychological lens by answering two questions.

1) *What* are the psychological *concepts* or *features* (i.e., root causes) that make humans susceptible to SE attacks? This corresponds to *what to exploit* from an attacker's point of view.
2) *What strategies* have been used by SE attacks? This corresponds to *how to exploit* what can be exploited from an attacker's point of view. To guide the

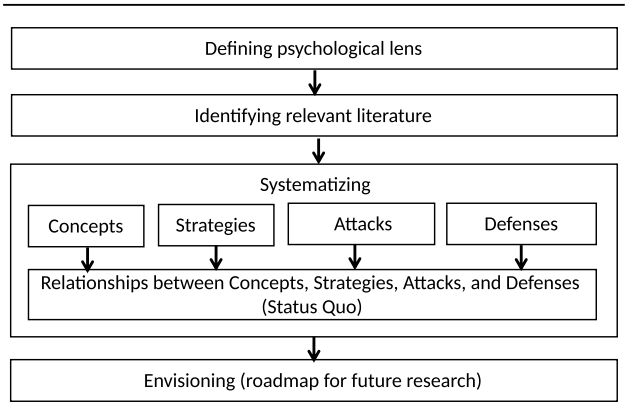


Fig. 1. ***General methodology.***

development of psychological lenses, we define their desired properties that are proposed here for the first time. Specifically, we say a lens is *competent* if it has the following properties.

a) *Robustness:* The psychological concepts associated with the psychological lens must be robust as supported by psychological studies.
b) *Relevance:* The psychological concepts associated with a psychological lens must be relevant to SE attacks.
c) *Comprehensiveness:* The psychological concepts associated with a psychological lens must be as comprehensive as possible, with respect to SE attacks.

At a high level, the robustness property would require one to look for concepts from the psychology literature, and the relevance and comprehensiveness properties would require one to have a deep understanding of the cybersecurity literature related to SE attacks.

Second, it would be ideal that there are competent psychological lenses that can be leveraged to identify the related literature and systematize the knowledge. However, we are not aware of any such competent lens in the literature. This means that researchers would have to define their own lens(es), which is true in this study. Under such circumstances, we propose that a psychological lens may be iteratively developed as follows: an initial, but not comprehensive, psychological lens is defined and leveraged to identify the related literature, from which other psychological concepts and strategies, as per the preceding 1) and 2), can be identified and systematized to define a competent lens. This is the approach used in the rest of the study, but other approaches may be possible.

Third, having defined a competent lens and identified the related literature, the systematization answers the following questions: 3) *what* psychological concepts have been more exploited than others by SE attackers? 4) *what* strategies have been more used than others by SE attackers? 5) *what* SE attacks have been reported in the literature? 6) *what* defenses have been proposed in the literature? and 7) *what* are the relationships between the psychological concepts, the strategies, the SE attacks, and the SE defenses?

Fourth, having systematized the relationships between the psychological concepts, strategies, SE attacks, and SE defenses, it is important to answer: 8) *what* are promising future research directions?

Instantiating the General Methodology via a Specific Lens: A systematization is based on a specific lens. In the rest of the study, we, respectively, instantiate the psychological concept and strategies in the general methodology as PFs and PTs, answering questions 1) and 2) mentioned above. Moreover, we have to define a psychological lens iteratively because there is no well-defined psychological lens. This allows us to address questions 3) and 4). Regarding question 5), we focus on attack *objectives*, namely, the motives of SE attackers, and *types*, for which we consider three major classes of SE attacks based on the media that they exploit—*email*, *website*, or OSN. For question 6), we focus on the defenses that leverage PFs and PTs. Regarding question 7), we systematize the PFs and PTs that are exploited by each SE attack and the defenses based on the PFs and PTs that are leveraged by them and the attacks that they defend against. The resulting understanding can guide us to propose a systematic research roadmap, answering 8).

III. DEFINING A SPECIFIC PSYCHOLOGICAL LENS

In this study, we propose using a specific lens, which is centered on the following notions of PF and PT.

Definition 1 (PF and PT): A PF is a human psychological characteristic or attribute that can be exploited by SE attacks. A PT is a strategy (i.e., method or approach) by which SE attackers exploit some PF(s) to encourage individuals to comply with their SE attacks.

Since we are not aware of any robust, relevant, and comprehensive psychological lens, we need to develop one under the guidance of the properties mentioned above. For this purpose, we propose defining a lens by considering initial PFs and PTs and then use them to identify other PFs and PTs.

Defining Initial PFs and PTs: For defining initial PFs, we propose leveraging the Big Five Personality Traits (BFPTs) and Cialdini's Principles of Persuasion [51] to define 11 initial PFs.

First, Cialdini's six Principles of Persuasion [51] are given as follows:

1) LIKING (or SIMILARITY), which refers to that one would be easily influenced by individuals one likes or individuals with common beliefs;
2) RECIPROCITY (or RECIPROCATION), which refers to that one may feel obliged to return a favor;
3) SOCIAL PROOF (or CONFORMITY), which refers to that one would imitate the behaviors of others;
4) CONSISTENCY (or COMMITMENT), which refers to the consistency of behavior or sticking to a promise;
5) AUTHORITY, which refers to obeying experts or orders from one's superior or authoritative figures;
6) SCARCITY, which refers to placing more value on things that are in short supply.

These PFs are robust because they are derived from field studies in the context of sales and marketing and are widely used. They are relevant because persuasion is widely used in SE attacks [52], [53], [54], [55], [56], [57], [58]. For example, [56] shows that AUTHORITY, SCARCITY, and LIKING are widely exploited by SE attacks; [59] shows that LIKING explains a person's tendency to fall victim to SE attacks; [60] shows that RECIPROCATION is the third most used principle of persuasion in SE attacks; [61] shows that the relevance of SOCIAL PROOF as evidenced by that Facebook users with ten or more friends tends to update their security settings according to what

their friends do; [56], [62], and [63] show the relevance of CONSISTENCY via individuals' dogmatic adherence to past decisions when making new decisions, which can be exploited by SE attacks; [54], [56], and [64] show that AUTHORITY has been widely exploited by SE attacks; and [56] shows that SCARCITY is often exploited by SE attacks perhaps because people care about what they may lose or miss (e.g., money, goods, or services).

Second, we use BFPT [65], [66], [67] as initial PFs:

1) OPENNESS, which refers to one's active imagination and insight toward new ideas or objects;
2) CONSCIENTIOUSNESS, which refers to one's thoughtfulness, impulse control, and goal-directed behaviors;
3) EXTRAVERSION, which refers to the degree to which one is sociable, assertive, talkative, and emotionally expressive;
4) AGREEABLENESS, which refers to one's attributes in relation to trust, altruism, kindness, affection, and other prosocial behaviors;
5) NEUROTICISM, which refers to one's moodiness and emotional instability.

These PFs are *robust* because they constitute the basic structure of human personality [66], they are good indicators of behavior [67], they have been studied in different languages and cultures over decades [68], [69], [70], and they are relatively stable across the lifespan and can help predict life outcomes ranging from career success to likelihood of divorce to lifespan longevity [71]. Note that they even appear to be present in other species [72]. Furthermore, they are relevant because personality traits are indicators of humans' susceptibility to SE attacks [34], [45]. For example, studies show that individuals with high AGREEABLENESS and EXTROVERSION are more susceptible to SE attacks [34], and individuals with high NEUROTICISM are less susceptible to SE attacks [45].

For defining initial PTs, we propose leveraging the following three PTs specified in [44], [54], [59], [62], [73], and [74], with which we are familiar, as initial PTs.

1) *Persuasion:* This PT is a natural choice because SE attackers often attempt to convince users to comply with their intent [44], [54].
2) *Pretexting:* This PT is a natural choice because it is widely used in various SE attacks [44], [73], [74].
3) *Impersonation:* This PT is the natural choice because SE attackers often impersonate legitimate users when sending malicious emails [59], [62].

These PTs have been used by attackers to exploit the PFs of victims, which will be elaborated on later.

Leveraging the Initial PFs and PTs to Identify Others: The preceding initial PFs and PTs serve as a starting point for identifying, extracting, and systematizing the other PFs and PTs from the references that will result from the literature search step. The resulting PFs and PTs can formulate a robust, relevant, and comprehensive psychological lens, as evidenced by the fact that the 11 initial PFs will lead to the identification of 46 PFs and the three initial PTs will lead to 16 PTs, which effectively address the comprehensiveness property. The basic idea is that when a reference investigates one PF or PT, the reference may also discuss other PFs and/or PTs.

IV. IDENTIFYING RELEVANT LITERATURE

To be as *comprehensive* as feasible, we identify other PFs and PTs than the initial ones by identifying and analyzing the relevant literature in three steps: selecting venues, determining search, and search results filtering.

A. Literature Venue Selection

Our domain knowledge suggests that the literature on SE attacks and defenses goes much beyond the traditional cybersecurity venues because of their interdisciplinary nature. This prompts us to consider a range of digital libraries, including: IEEE (including Symposium on Security and Privacy, European Symposium on Security and Privacy, and IEEE TRANSACTIONS), ACM (including ACM Conference on Computer and Communications Security, ACM ASIA Conference on Computer and Communications Security, Annual Computer Security Applications Conference, and ACM Transactions), Usenix (including Usenix Security Symposium), ISOC (including Network and Distributed System Security Symposium), Elsevier, Springer (including European Symposium on Research in Computer Security and Detection of Intrusions and Malware & Vulnerability Assessment), PlosOne, Wiley, *Frontiers in Psychology*, and *Information and Computer Security*.

Although our focus is on academic studies, as they aim at principled investigations, we also consider the so-called "grey" literature [75], namely, nonpeer-reviewed sources, which can be useful because SE attacks may be reported in online media but have not been investigated in academic literature.

B. Literature Search Method

Owing to the evolution of SE attacks, we propose considering literature published in the last ten years, which also helps reduce the amount of literature. Then, we conduct a keyword-based literature search as follows.

On the one hand, we search academic literature as follows. First, we conduct the initial search for academic literature in the digital libraries mentioned above as follows. 1) to identify survey literature, we use keywords [social engineering attacks and (survey or literature review)] to search in each digital library; 2) to identify research literature with respect to PFs and PTs, we search each digital library using the 11 initial PFs and the three initial PTs as keywords, respectively; 3) to identify SE attack literature, we use the following keywords: phishing, vishing, smishing, business email compromise, [(social engineering attacks) and taxonomy], and types of social engineering attacks; and 4) to identify SE defense literature, we use the following keywords: (social engineering attacks) and defenses or (countermeasures or prevention).

Second, we iterate the preceding step by using the keywords that are newly identified from the search results obtained in the preceding iteration. These new keywords are divided into four categories: PF, PT, attack, and defense, which are identified based on our examination of the returned literature. The newly identified PF keywords (as candidate PFs, which will be scrutinized later) include *cognitive miser*, *overconfidence*, *loneliness*, *hopelessness*, *disobedience*, *perceptual contrast*, *self-efficacy*, and *subjective norm*. The newly identified PT keywords (as candidate PTs, which will be scrutinized later) include *priming*, *quid pro quo*, *foot in the door*, *decoy effect*, and *loss aversion*. The newly identified attack keywords include *ad fraud*, *app spoofing*, and *QRishing*. The iteration halts when we do not see new terms that deserve to be considered as keywords, which is subjective.

Third, we still encounter the situation that we are aware of some psychological attributes (i.e., *absentmindedness*, *freewheeling*, and *hopelessness*) and some attacks (e.g., *honey trap* [76] and *angler phishing* [77]). They are clearly relevant but hit no academic literature in those digital libraries, perhaps because their investigation is not published in the venues mentioned above. Thus, we further use Google Scholar with these keywords to conduct extended searches that go beyond the digital libraries mentioned above. This leads to the accommodation of [43], [48], [78], [79], [80], and [81] on *absentmindedness*, [80] on *freewheeling*, [82] and [83] on *hopelessness*, [76] on honey trap, and [84] on angler phishing.

On the other hand, we use Google to search for gray literature via phrases including (the latest cyber security attacks), (the most cybersecurity attacks), and (security breaches using social engineering attacks). Moreover, we use reports from the APWG, which analyses trends of SE attacks. This leads to 11 references: [1], [2], [3], [85], [86], [87], [88], [89], and [90].

C. Search Results Filtering

The search leads to 752 papers. We manually examine each paper based on its technical relevance to our study. We eliminate the papers that only mention some terms (i.e., search keywords) without presenting substantial investigation; 286 papers fall into this category. We further eliminate the ones that do not consider PFs or Internet-based SE attacks (e.g., studies considering tailgating in the physical worlds but not cyber); 184 papers fall into this category. We further eliminate the ones that do not present a quality exploration (e.g., offering no significant new understanding); 48 papers fall into this category. In total, the filtering process led to 234 papers left. Together with ten papers suggested by reviewers, we have 244 papers for systematization, including 38 survey/review papers and one handbook [47]. Note that the other cited references are for purposes such as analyzing real-world attacks (e.g., the industrial report on Colonial Pipeline attack [87]) and relating the envisioned future research to other endeavors.

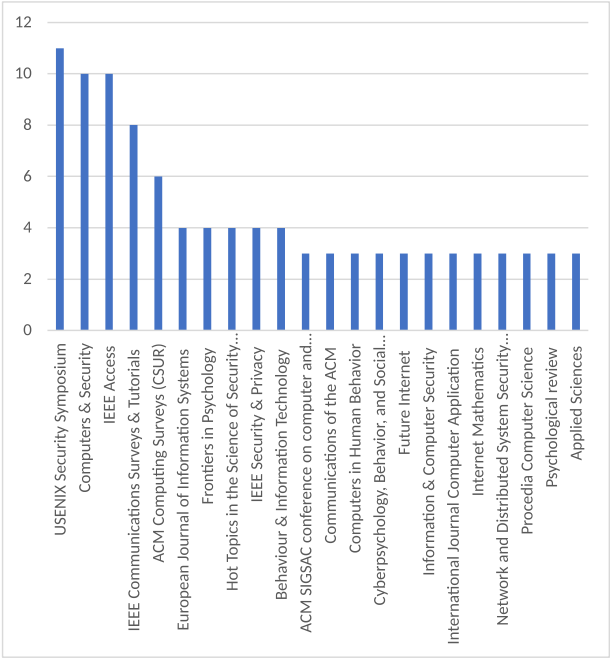


Fig. 2. ***Venues publishing at least three papers cited in this study (e.g., 11 publications in the Usenix Security Symposium).***

D. Venue Analysis

As a side-product, we analyze the academic venues that publish SE attack and/or defense papers, which would be useful for researchers (readers) to identify the relevant venues for publication (studying). Fig. 2 shows the venues that publish at least three references cited in this article. We observe that Usenix Security Symposium, *Computer & Security*, IEEE Access, and IEEE Communications Surveys and Tutorials (IEEE CST) publish most papers (i.e., 96 papers in total). However, we also observe that 149 papers are published in 123 distinct venues (by publishers including IEEE, Springer, Wiley, ACM, and ISOC), and 11 publications come from psychological venues (e.g., *American Psychologist*, *Annual Review of Psychology*, and *Cognitive Psychology*).

V. SYSTEMATIZING PFs

Now, we describe how we identify other PFs (other than the 11 initial PFs) from the references and how we categorize them.

A. Identifying Other PFs

The 11 initial PFs would not be comprehensive, suggesting us to identify and extract other PFs from the references as follows. First, if a paper explicitly states that a psychological attribute has an impact on SE attacks, we make the attribute a candidate PF (e.g., *workload*, *stress*, *vigilance*, and *expertise* from [44]). Second, we use the initial and newly identified candidate PFs as keywords to search the references via Adobe Reader Advanced Search to identify other candidate PFs. This is possible because a

paper mentioning one PF may contain other psychological attributes or candidate PFs. We repeat this step until we cannot identify new candidate PFs. This process leads to 42 candidate PFs, or 53 in total (including the 11 initial PFs), which, however, contain some redundancy. This prompts us to eliminate the redundant ones while using the 11 initial PFs as the baseline. Details are given in the following.

First, any candidate PF that is redundant with any of the 11 initial PFs is eliminated. We encounter four such candidate PFs: *similarity* is redundant to LIKING, *commitment* is redundant to CONSISTENCY, *reciprocity* is redundant to RECIPROCATION, and *herd mentality* is redundant to SOCIAL PROOF. This leads to 49 candidate PFs.

Second, if two candidate PFs are similar in their psychological meaning, we keep the candidate PF that is investigated in a quantitative fashion in this article from which it is extracted because a quantitative study would represent a deeper understanding than a qualitative study. We encounter that two candidate PFs: *false consensus effect* (extracted from [91]) and *social proof* (extracted from [61]) are similar, and we keep the latter because it is studied in a more quantitative fashion in [61]. This leads to 48 candidate PFs remaining.

Third, if two candidate PFs are similar, but none of them are investigated in a quantitative fashion, we keep the PF that is more relevant to this article. We encounter two such candidate PFs: *freewheeling* (extracted from [80]) and *freeloader* (extracted from [92]); we keep the former and use it to represent both. This leads to 47 candidate PFs remaining.

Fourth, if two PFs are considered redundant and both are investigated in a quantitative fashion, we keep the one that is more often used in the literature according to our domain knowledge in psychology. We encounter two such PFs: *inattentiveness* (extracted from [93]) and *lack of vigilance* (extracted from [94]) are redundant because they essentially represent the same psychological attribute. Since they are both investigated in a quantitative fashion, we keep the former because it is used in the literature more often (perhaps because of its succinctness). This leads to 46 candidate PFs remaining.

In order to systematize the 46 PFs, we leverage our domain knowledge in psychology to classify them into five classes.

1) *Social psychology* PFs, which describe individuals' interpersonal attributes. The six initial PFs from Cialdini's Principles of Persuasion belong to this class.
2) *PID* PFs, which are individuals' relatively stable attributes and personality traits. The five initial PFs from the BFPT belong to this class.
3) *Cognitive* PFs, which describe how individuals process information.
4) *Emotion* PFs, which describe individuals' feelings, motivational states, approaches, and avoidance behaviors.
5) *Workplace* PFs, which describe cultural and organizational interactions in a workplace.

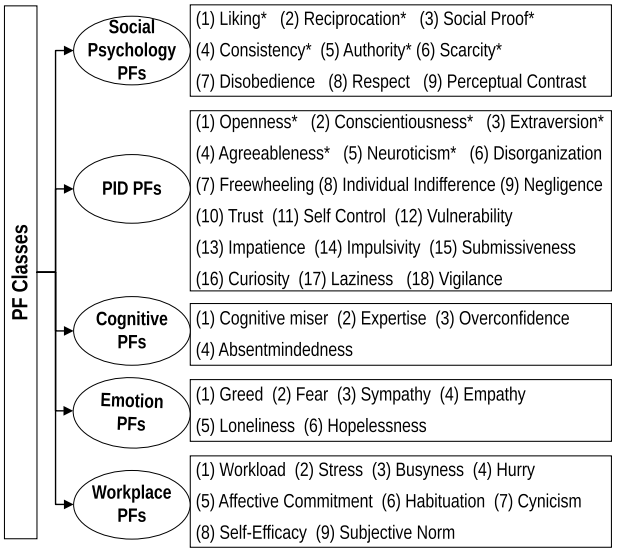


Fig. 3. ***Summary of the 46 PFs in five classes, where the 11 initial PFs are indicated by "*" (including the six initial social psychology PFs from Cialdini's Principles of Persuasion and the five initial PID PFs from BFPT.***

Fig. 3 highlights the five classes of 46 PFs. We use the preceding categorization because: 1) it is in line with the traditional branches and subdivisions in psychology and 2) it has potential utility, such as helping defenders to design effective defenses (e.g., the PFs that should be the focus of a training scheme would depend on the trainees' job duties). Nevertheless, researchers who do not prefer this classification can categorize the 46 PFs according to their needs. We admit that our classification is subjective because: 1) it is based on our domain knowledge and 2) several PFs can fall into multiple classes, and in such cases, we assign them to the ones we believe appropriate. We note that other categorizations are possible. This should not be treated as a weakness of this study because multiple categorizations might lead to different findings, which can be synergized to formulate a holistic body of knowledge. To illustrate the preceding 2), we mention three examples: 1) we treat OVERCONFIDENCE as a *cognitive* PF even though it could be considered a PID PF because it refers to stable individual differences; 2) we treat FEAR as an *emotion* PF because it is more an emotional factor than a social psychology factor despite that it is seemingly very related to AUTHORITY (i.e., a *social psychology* PF); and 3) we treat STRESS and WORKLOAD as workplace factors, while they affect cognition and can of course also be social in nature.

B. Systematizing Social Psychology PFs

These PFs describe one's interpersonal behaviors and often involve connection, influence, and demand/request

interactions between the individual and others. For example, people have a natural tendency to obey authorities, which may be reinforced via societal training [58]. This can be exploited by attackers to craft messages to exploit people's obedience. This class has nine PFs, among which the first six are derived from Cialdini's Principles of Persuasion.

1) *Liking (or similarity):* This initial PF describes individuals' tendency to react positively to those with whom they have a relationship [95]. It reflects that people may be persuaded to obey others if they display certain favorable or familiar characteristics [65]. It has been exploited to create profiles that portray trusted traits or appear friendly to lure victims [96]. One study [60] shows that LIKING is exploited by 91% of the SE attacks investigated. Another study [59] shows that LIKING is an individual variable that explains a person's tendency to fall victim to SE attacks.
2) *Reciprocity (reciprocation):* This initial PF describes humans' tendency to pay back a favor [52], [88], [97]. This is part of human nature as one often feels indebted to the person who helped one earlier, even if the requested payback is not of the same magnitude as the one that was received. This puts the person demanding the payback in an advantageous position. One study [60] shows that RECIPROCATION is the third most used principle of persuasion exploited by SE attacks, only after AUTHORITY and LIKING.
3) *Social proof (or conformity):* This initial PF describes humans' tendency to imitate others regardless of the importance or correctness of the behavior [56], [62], [65], [98], [99]. It can put people at risk because they tend to let down their guard when everyone else appears to share the same or a similar behavior [95]. One study [61] with 50 000 Facebook users shows that users with ten or more Facebook friends tend to update their security settings after being informed that their friends have updated security settings.
4) *Consistency (or commitment):* This initial PF describes the degree to which one is dedicated to a person, object, task, or ideal [99]. SE attacks exploit it to persuade their victims [97]. One study [62] shows that dogmatic adherence to past decisions may influence one's decisions in the future, which can be exploited to wage attacks [56], [63].
5) *Authority:* This initial PF describes power or dominance over someone [100]. SE attacks use AUTHORITY to lure their victims to divulge confidential information, especially through spear phishing [101]. One study [60] shows that, out of the six principles of persuasion, the effect of this PF alone exceeds that of the other five principles together.
   Another study [102] with 612 participants shows that individuals who are more obedient to authority are more susceptible to SE attacks.
6) *Scarcity:* This initial PF describes a lack of goods/services. It is widely exploited in online scams [60], [96] and phishing emails [56]. It is often exploited together with the AUTHORITY PF to lure victims into submitting to their demands [101], [103].
7) *Disobedience:* This PF, extracted from [22], [80], and [104], describes one's dogmatic refusal to obey authority or rules, making one susceptible to SE attacks [80]. While it is well known that people who are more trusting and obedient to authority are more susceptible to SE attacks [22], it is less known that employees' willful disobedience can also be exploited by SE attacks [104].
8) *Respect:* This PF, extracted from [62], [97], [105], and [106], describes one's esteem for another and reflects the degree to which the other is perceived as valuable or worthwhile [62]. For example, an individual may not question a suspicious request from a friend (e.g., an unsolicited email that contains a link) out of respect for their relationship [105]. This PF may be exploited together with AUTHORITY [97], [106].
9) *Perceptual contrast:* This PF, extracted from [107] and [108], describes a mental deception when comparing two items in succession, where the first one influences how the second one will be perceived [107], [108]. For example, one may consider a fake product as more valuable than a legit product of lower contrast.

Real-World Example: An excerpt from a real-world phishing email impersonating Coinbase is: *"Hi John, Financial regulations require us to confirm your info by October 01, 2022, or your account will be restricted."* This attack exploits the AUTHORITY PF by impersonating the cryptocurrency platform Coinbase and depicting an authority that people are generally trained to respect and act as authority demands.

What Insight Can We Draw? The systematization suggests that all nine social psychology PFs have been exploited by SE attacks perhaps because of humans' inherently social nature. However, there is a lack of quantitative understanding of their impact on SE attack effectiveness.

Insight 1: All nine social psychology PFs have been exploited by SE attacks, with AUTHORITY being exploited most often.

C. Systematizing PID PFs

These PFs describe the uniqueness of individuals in terms of their personality traits and stable attributes as related to mental abilities, vocational interests, religious beliefs, political attitudes, sexuality, and more [89]. For example, some people are habitually more meticulous and attentive to details than others, while others are habitually more trusting. These PFs are usually exploited by attackers to send a large number of malicious emails, hoping that recipients with these PFs will fall victim to them. There are 18 PID PFs in total, among which the first five are the initial PFs from the BFPT.

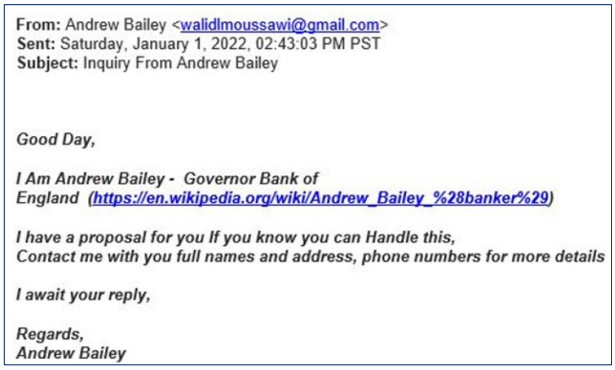
1) *Openness:* This initial PF describes one's active imagination and insight [90]. Individuals of high openness are often curious about the world and other people, eager to learn new things, enjoy new experiences, and are more adventurous and creative. High openness predicts high susceptibility to phishing attacks [28], [65].
2) *Conscientiousness:* This initial PF describes one's thoughtfulness, impulse control, and goal-directed behaviors. Highly conscientious people tend to be more organized, mindful of details, self-disciplined, goal-oriented, proficient planners, and considerate about how their behaviors may affect others [65], [90]. One study [109] shows that people high in conscientiousness are less susceptible to spear phishing attacks.
3) *Extraversion:* This initial PF, also known as EXTROVERSION, describes the degree to which one is sociable, assertive, talkative, and emotionally expressive [90]. People high in extraversion are outgoing and tend to gain energy in social situations. One study [110] shows that EXTRAVERSION (and OPENNESS and AGREEABLENESS) increase one's susceptibility to phishing emails.
4) *Agreeableness:* This initial PF describes one's attributes related to trust, altruism, kindness, affection, and other prosocial behaviors [90]. One study [111] shows that people high in AGREEABLENESS are more susceptible to phishing attacks.
5) *Neuroticism:* This initial PF describes one's moodiness and emotional instability. People of high NEUROTICISM often exhibit mood swings, anxiety, irritability, and sadness [90]. Individuals high in NEUROTICISM are more susceptible to phishing attacks [111].
6) *Disorganization:* This PF, extracted from [80], describes the tendency of an individual to act without prior planning or to allow their environment to become or remain unstructured or messy. These conditions may blind them to anomalies or cues of attacks, resulting in a higher susceptibility to SE attacks [80].
7) *Freewheeling:* This PF, extracted from [80], describes the degree to which one disregards rules or conventions and the degree of one's unconstraint or disinhibition, which contributes to one's susceptibility to SE attacks. One report [112] suggests that cybercriminals can freewheel to innovate attacks, but defenders do not have that freedom due to company bureaucracy and policies, meaning that defenders who try to freewheel may end up exposing the company to cyberattacks.
8) *Individual indifference:* This PF, extracted from [81], describes the degree to which one shows disinterest toward a task. One study [81] shows that a sustained indifference toward security can cultivate a culture of risky behaviors, which can be exploited by SE attacks. The study also shows that there is a degree of perceived importance of cybersecurity in organizations as evidenced by employees, including management and sometimes security staff, exhibiting indifference toward cybersecurity policies and procedures.
9) *Negligence:* This PF, extracted from [113], [114], [115], and [116], describes one's failure to take proper care of a particular task, causing security breaches. One study [115] reports that 27% of data breaches are due to negligent employees or contractors, who usually have remote access to organizations' internal networks. Other studies [113], [114], [116] show that negligence is the chief reason that users fall victim to phishing attacks.
10) *Trust:* This PF, extracted from [22], [59], [100], [102], and [117], describes the tendency of humans to trust or believe in others. People who are more trusting are more susceptible to SE attacks [22], which is not surprising because developing trust is a key element of SE attacks [100], [117]. Moreover, people who are predisposed to trust others that they view as likable are more likely to fall victim to scams [59]. A study [102] with 612 participants shows that people who are more trusting succumb more frequently to SE attacks.
11) *Self-control:* This PF, extracted from [81], [96], [118], and [119], describes one's ability to regulate one's decision-making processes in the face of strong emotions and desires. A lack of self-control allows individuals to fall victim to online scammers [96]. Individuals with low self-control tend to exhibit a higher willingness to take risks in SE attack situations [81], [118], [119].
12) *Vulnerability:* This PF, extracted from [120], describes the degree to which one is in need of special care, support, or protection because of age, disability, or risk of abuse or neglect. The study [120] aimed to identify those at greater risk of falling victim to SE attacks in an organization and showed that employees with one year of service or less are more susceptible to spear phishing (52.07%) than employees with eight years of service (23.19%).
13) *Impatience:* This PF, extracted from [119], describes one's frustration while waiting for a particular event to occur or at the length of time needed to accomplish a task. Impatient individuals may be more susceptible to SE attacks because they do not carefully examine the contents or cues of SE attacks, especially when they focus on immediate gratification [119].
14) *Impulsivity:* This PF, extracted from [28], [121], and [122], describes some humans' tendency to act without much forethought [28]. A study [121] with 53 undergraduate students showed that participants low in impulsivity are less susceptible to phishing. Another study [122] shows that individuals who are high in sensation-seeking, a form of impulsivity, are more likely to be vulnerable to scams.

15) *Submissiveness:* This PF, extracted from [123], describes the degree of one's readiness to conform to the authority or will of others. The study [123] with approximately 200 participants finds that high submissiveness implies a high susceptibility to phishing emails.
16) *Curiosity:* This PF, extracted from [96], [124], and [125], describes the degree to which one desires to know something. Online scammers exploit victims' curiosity to encourage errors in judgment and decision-making [96] or to serve as a persuasion technique to lure their victims [124], [125].
17) *Laziness:* This PF, extracted from [126], describes the degree of one's voluntary inability to carry out a task with the energy required to accomplish it. The study [126] shows that laziness makes people unwilling to do the necessary work or apply the effort to mitigate risk and, thus, can make them more susceptible to SE attacks.
18) *Vigilance:* This PF, extracted from [94], [116], and [127], describes the degree to which one is watchful for possible dangers or anomalies. High vigilance makes one less susceptible to SE attacks [94], [116]. A phishing experiment with 3000 university students finds that VIGILANCE reduces susceptibility to scams [94].

Real-World Example: One real-world example is the exploitation of the NEGLIGENCE PF in the Colonial Pipeline attack. The attacker, known as DarkSide, hacks into the Pipeline's computer network via a compromised VPN password of an old VPN account that was no longer in use [87]. This attack can be attributed to the NEGLIGENCE PF of the defender for not revoking the VPN account that is no longer in use.

Fig. 4 shows another real-world email that exploits PID PFs. The email exploits: 1) the CURIOSITY PF, which is effective when the recipient has a high desire to know the governor of the Bank of England and 2) the IMPULSIVITY PF, which is effective when the recipient is acting without much forethought. Moreover, the legitimate link exploits: 1) the lack of the VIGILANCE PF that is effective when the recipient is not watchful and 2) the AGREEABLENESS PF that is effective when the recipient easily trusts strangers. Nevertheless, the red flags include the **From:** field, which is a Gmail account rather than a Bank of England email, and the **Subject:** field, which has the name of the Governor of the Bank of England, Andrew Bailey. These two red flags require the due diligence of recipients in dealing with the email.

What Insight Can We Draw? PID PFs are diverse and have all been exploited by SE attacks. The systematization above suggests that individuals high in OPENNESS, EXTRAVERSION, AGREEABLENESS, NEUROTICISM, DISORGANIZATION, FREEWHEELING, INDIVIDUAL INDIFFERENCE, NEGLIGENCE, TRUST, VULNERABILITY, IMPATIENCE, IMPULSIVITY, SUBMISSIVENESS, CURIOSITY, or LAZINESS are more susceptible to SE attacks and individuals low in CONSCIENTIOUSNESS and SELF-CONTROL are more susceptible to SE attacks. However, there is a lack of quantitative characterization of the impact of these PFs. Moreover, there are intriguing issues that are yet to be understood. For example, even though individuals with a high VIGILANCE usually would not open a suspicious email, they may end up opening it due to spontaneous CURIOSITY [127]. This highlights the possible interactions between PFs, namely, that one PF may dominate another under certain circumstances. Similarly, even users practicing zero trust can fall prey to NEGLIGENCE and open a malicious email.


From: Andrew Bailey <walidlmoussawi@gmail.com>
Sent: Saturday, January 1, 2022, 02:43:03 PM PST
Subject: Inquiry From Andrew Bailey

Good Day,

I Am Andrew Bailey - Governor Bank of England (https://en.wikipedia.org/wiki/Andrew_Bailey_%28banker%29)

I have a proposal for you If you know you can Handle this, Contact me with you full names and address, phone numbers for more details

I await your reply,

Regards,
Andrew Bailey

Fig. 4. ***Real-world malicious email with a legitimate Wikipedia link about the governor of the Bank of England, showing how PID PFs are exploited.***

Insight 2: SE attacks have exploited all the 18 PID PFs, but TRUST and NEGLIGENCE are the most exploited.

D. Systematizing Cognitive PFs

These PFs describe how an individual processes information, including heuristics that they may use, the knowledge that they may possess, their degree of confidence, and the attention that they may give. Key aspects of cognition include attention, memory, and knowledge. There are four cognitive PFs.

1) *Cognitive miser:* This PF, extracted from [128], [129], and [130], describes one's use of decision-making heuristics, namely, the use of mental shortcuts in a decision-making process. People sometimes behave as COGNITIVE MISERS and rely on heuristic-based processing to make decisions [129]. One study [130] argues that people are motivated tacticians and will apply a COGNITIVE MISER (or naive scientist) approach based on the urgency, perceived importance, and complexity of the situation. Another study [130] shows that using COGNITIVE MISER is faster and demands less cognitive effort but is error-prone.
2) *Expertise:* This PF, extracted from [97], [131], [132], [133], and [134], describes one's knowledge about a particular domain. One study [131] shows that EXPERTISE plays a role in raising an individual's perception of risk associated with OSNs, but the

perceived risk does not significantly increase individuals' competence in coping with these threats. Qualitatively speaking, EXPERTISE does not necessarily make one less susceptible to SE attacks [97], [132]; quantitatively speaking, EXPERTISE, when effective, can make one able to cope with SE attacks (i.e., incurring lower false-positive and false-negative rates [133]). Another study [134] shows that the EXPERTISE associated with a given social-demographic background may affect the prioritization of advice in coping with online threats.

3) *Overconfidence:* This PF, extracted from [96], [98], [135], [136], and [137], describes humans' tendency in having too much confidence in themselves [96], especially their ability in detecting phishing [135], but this can be improved via education and training [98]. This PF may correlate with self-confidence, a PID factor [136]. One study [137] conducted an experiment with 53 undergraduate students (34% computer science majors and 66% psychology majors), showing that approximately 92% of the participants misclassified phishing emails even though 89% had indicated earlier that they were confident of their ability to identify phishing emails.
4) *Absentmindedness:* This PF, extracted from [43], [48], [78], [79], [80], and [81], describes the degree to which one's attention is diverted from a task. One study [78] shows that employees' ABSENTMINDEDNESS is positively related to emotional exhaustion, which negatively affects one's job performance; [79] and [80] show that absentminded people might click phishing links because they do not pay attention to what they are doing; [43], [48], and [81] show that participants may not even notice or check system warnings such as "Warning: This email is from an external source."

Real-World Example: Fig. 5 is a real-world email showing how the attacker exploits *cognitive* PFs, where we redact the recipient information for privacy reasons. We observe that the only visible give-away is the **From:** field, which is highlighted to show that the sender is not Chase bank as it claims. If a recipient is in a high ABSENTMINDEDNESS state when opening the email, the recipient may click on the link in the email. The email also exploits the COGNITIVE MISER PF when the recipient spends less mental effort on the email, especially when deceived by the Chase logo and brand.

What Insight Can We Draw? Cognitive PFs likely play important roles in influencing individuals' susceptibility to SE attacks. The systematization above suggests: 1) humans tend to use mental shortcuts in their decision-making; 2) EXPERTISE can, but does not always, reduce one's susceptibility to SE attacks; 3) people tend to be overconfident in dealing with SE attacks; and 4) ABSENTMINDEDNESS increases susceptibility to SE attacks. However, there is a lack of quantitative understanding of the impact of these PFs on users' susceptibility, which remains an open problem.

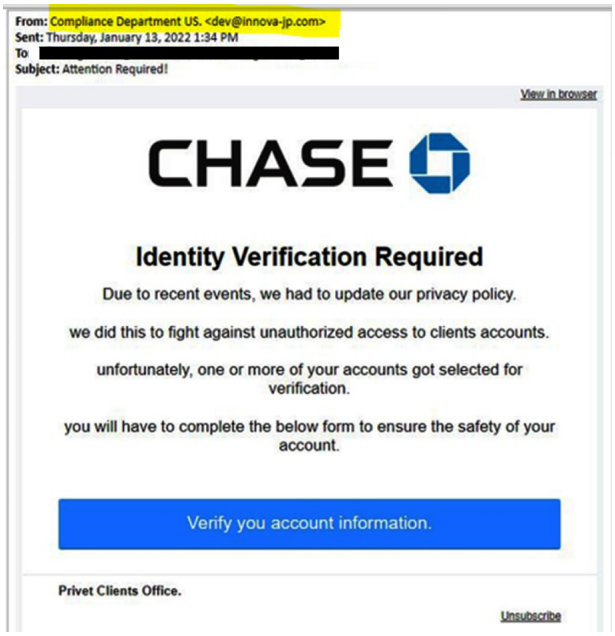


Fig. 5. ***Real-world phishing email impersonating Chase bank. The email content exploits two cognitive PFs in* cognitive miser *and* absentmindedness.**

Insight 3: Humans are susceptible to SE attacks partly because they use shortcuts in reasoning and are often overconfident and absentminded, and because expertise may not be as helpful as desired. Among the four cognitive PFs, the COGNITIVE MISER PF is the most exploited because humans often use mental shortcuts in their reasoning.

E. Systematizing Emotion PFs

These PFs describe human feelings, motivational states, and approach or avoidance behaviors. There are six emotion PFs.

1) *Greed:* This PF, extracted from [20], [29], [32], [63], [138], [139], and [140], describes one's intense desire for something, especially wealth, power, or food. GREED is often exploited in phishing emails [124] and is often paired with need (i.e., the attacker knows what a victim needs and thus presents what the victim needs as bait) [64]. GREED is recognized by some researchers as a human limitation when comparing human-based security versus technology-based security [5], [114]. Moreover, the greedier a person is, the more likely the person will fall victim to SE attacks [115].
2) *Fear:* This PF, extracted from [32], [62], [94], [124], [141], [142], and [143], describes one's belief that something painful, dangerous, or threatening may happen. It is relevant because situations that evoke FEAR incur a strong avoidance reaction in both behavioral responses and cognitive processing [143]. One

study [62] shows that SE attacks are effective against people who feel FEAR toward influential people. Another study [124] treats FEAR as one of the phishing persuasion techniques. We treat it as a PF because it is a universal human emotion.

3) *Sympathy:* This PF, extracted from [63], [96], [99], [144], and [145], describes the emotional state of individuals who understand the mental or emotional state of another person without necessarily feeling the same emotion. It can make humans susceptible to SE attacks [99] as attackers often seek to gain people's SYMPATHY [63], [144], [145].
4) *Empathy:* This PF, extracted from [96], [106], [124], [146], [147], [148], and [149], describes the emotional state of an individual who personally relates to the mental or emotional state of another person based on past experiences with the same state. Scammers often exploit EMPATHY as an intuitive behavior [106] or a persuasion technique [124] to achieve their goals [96], [106].
5) *Loneliness:* This PF, extracted from [52], [96], [150], [151], and [152], describes one's subjective perception of discrepancy between one's desire and one's actual social companionship, connectedness, or intimacy. It is often exploited by attacks because the feeling of alienation from peers makes people susceptible [96], [151]. The psychological reason is that attackers can exploit the need for attention that accompanies the feeling of loneliness, which may be even more relevant to elderly people [52]. One study [152] with 299 participants finds that loneliness positively predicts problematic Internet uses that can be exploited by SE attacks.
6) *Hopelessness:* This PF, extracted from [82] and [83], describes an individual's mental state in despair of lack of hope, or the feeling that things cannot be improved, or the despair of not being able to redress their grievances at the workplace. People in this state are easily deceived by attacks that offer false hope.

Real-World Example: Here is an example from a real-world scam email. *"Good morning, I need a favor from you. I need to get a Google Play card for my niece who is sick. It's her birthday today and I cannot do this now, because I'm currently traveling for a two days trip. I tried purchasing it online, but it is not going through. Please can you help me get them from any store around you? I'll repay you when I get back."* This attack attempts to trigger emotion PFs, especially SYMPATHY and EMPATHY, from a recipient.

What Insight Can We Draw? The preceding systematization suggests that *emotion* PFs have been widely exploited by SE attacks, and GREED, FEAR, SYMPATHY, EMPATHY, LONELINESS, and HOPELESSNESS all make humans susceptible to SE attacks. However, there is no quantitative understanding of their impact on individuals' susceptibility to SE attacks.

Insight 4: All the six emotion PFs have been exploited by SE attacks, but FEAR and GREED appear to be the most targeted.

F. Systematizing Workplace PFs

These PFs have to do with the culture and organizational structure of the workplace. This is relevant because various workplace environments may result in various levels of stress, employee engagement, or employee loyalty. Attackers can exploit what happens at workplaces to attack people. There are nine workforce PFs.

1) *Workload:* This PF, extracted from [127] and [141], describes the amount of work that one has to do. A survey of 488 employees at three hospitals shows that employee workload level is positively correlated with the likelihood of employees clicking on phishing links [127]. Another study finds that subjective mental workload creates memory deficits that lead to an inability to distinguish between real and fake messages, increasing susceptibility to attacks [141].
2) *Stress:* This PF, extracted from [96] and [141], describes the physical, emotional, or psychological strain placed on a person. Studies [96], [141] show that when people are stressed, their ability to notice suspicious communications (e.g., distinguishing real from fake messages) is reduced, making them more susceptible to SE attacks.
3) *Busyness:* This PF, extracted from [81] and [153], describes the degree to which one has too many things to do, which may or may not be associated with the workload. People with a high state of BUSYNESS are more susceptible to phishing emails, as they do not pay much attention to details [81] and may have reduced cognitive processing [153].
4) *Hurry:* This PF, extracted from [24] and [81], describes the degree that one is rushing to complete a task. Hurried people may not adhere to secure practices because they reduce the amount of time available for the individual's active task [81], causing them to be susceptible to SE attacks because they do not take the time to analyze the email with sufficient attention [24].
5) *Affective commitment:* This PF, extracted from [102], describes one's emotional attachment to an organization. One study [102] with 612 participants shows that people of high AFFECTIVE COMMITMENT are more likely to fall victim to SE attacks. For example, the love for one's organization may blind one in objective reasoning, as one may focus on satisfying the organization, which can be exploited by SE attacks [102].
6) *Habituation:* This PF, extracted from [154], describes one's tendency to perform a particular task repeatedly and get desensitized to it. The study [154] on how users perceive and respond to security messages using eye-tracking with 62 participants finds that people gazing less at warnings over successive viewings

(i.e., they are more habituated to warnings) are less attentive to security warnings. In other words, habituation increases susceptibility to attacks.

7) *Cynicism:* This PF, extracted from [155], describes one's tendency to willingly allow others to be harmed in order to get an advantage. One study [155] shows that having an unpleasant boss at work can lead to cynicism, especially in disgruntled employees, and this can be exploited by SE attacks.
8) *Self-efficacy:* This PF, extracted from [138], [156], [157], and [158], describes one's belief or self-confidence in producing an intended result. It is an important PF in human functioning [156] and a determinant in email-related behavior [138], [157], [158]. Lack of SELF-EFFICACY makes one susceptible to phishing [159].
9) *Subjective norm:* This PF, extracted from [113], [160], and [161], describes one's belief that an important person or group of people will approve or support a particular behavior, and this may lead one to behave in a particular way. This often causes social pressure in the workplace [160] and can make employees click on phishing emails [161]. This is related to how individuals worry about what other people think about them [113].

Real-World Example: There is an excerpt of a real-world malicious email. *"How are you doing? Are you available at the moment? I need your assistance to handle a little project. Can you please handle this for me on behalf of the organization?"* This message exploits the HABITUATION PF, which pertains to the routine of an employee to perform such tasks, and the AFFECTIVE COMMITMENT PF, attempting to make use of the individual's attachment and commitment to the organization.

What Insight Can We Draw? The preceding systematization suggests that workplace PFs have a significant impact on employees' susceptibility to SE attacks. Individuals of high WORKLOAD, STRESS, BUSYNESS, HURRY, AFFECTIVE COMMITMENT, HABITUATION, CYNICISM, or SUBJECTIVE NORM are more susceptible to SE attacks; individuals of low SELF-EFFICACY are more susceptible to SE attacks.

Insight 5: SE attacks have exploited all nine workforce PFs. Among all 46 PFs, AUTHORITY appears to be the most exploited one.

VI. SYSTEMATIZING PTs

A. Identifying Other PTs

We identify other PTs from the references in a fashion similar to the identification of candidate PFs. First, we examine the papers that contain any of the three initial PTs to identify other candidate PTs. Second, we use newly identified candidate PTs as keywords to search the papers to identify other candidate PTs in a recursive fashion. This process halts when it identifies no more candidate PTs, leading to 13 candidate PTs. We do not observe redundancy among the three initial PTs and the 13 candidate PTs, leading to 16 PTs in total.

B. Systematizing PTs

Unlike PFs, it is more challenging to categorize the 16 PTs. One may suggest categorizing them based on the attacks that exploit them, but this is not ideal because one PT can be exploited by multiple kinds of attacks. Similarly, it is not ideal to categorize PTs based on the PFs exploited by them because one PT can exploit multiple kinds of PFs. As a result, we simply list the PTs without categorizing them.

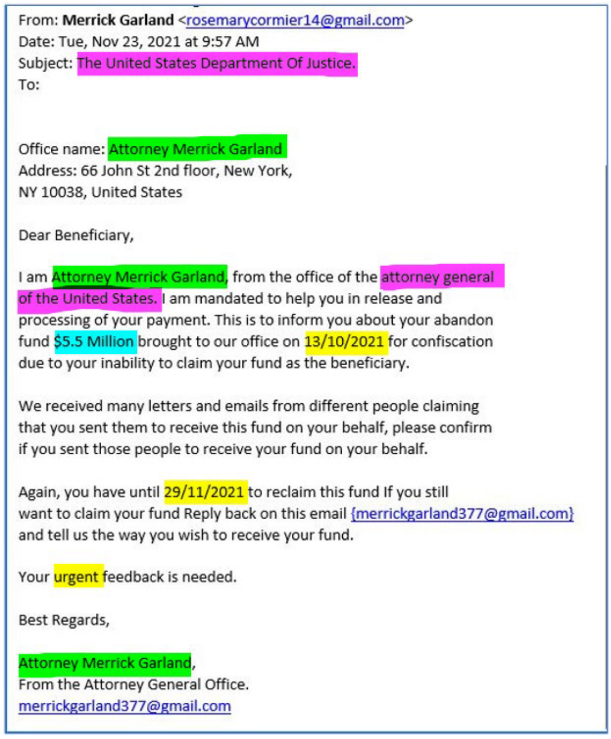
1) *Persuasion:* This initial PT encourages a particular behavior by exploiting the LIKING, RECIPROCATION, SOCIAL PROOF, CONSISTENCY, and AUTHORITY PFs. The effectiveness of this PT would depend on exactly which PFs are exploited and other factors such as age [52] and request type [162], [163]. This PT is widely used in email-based attacks such as phishing [54], [162], [164].
2) *Pretexting:* This initial PT increases the engagement of a victim with the attacker by exploiting the TRUST PF. For example, phishing emails often use this PT to increase responsiveness by adding elements that refer to current events such as holiday festivities or news [73], [74].
3) *Impersonation:* This initial PT assumes a false identity to increase a victim's compliance by exploiting the AUTHORITY, RESPECT, and TRUST PFs. In OSN-based attacks such as honey trap [76], attackers use fake profiles to lure victims into interacting with them [62]. For example, an attacker using BEC assumes the persona of an executive to ask a victim to transfer money to the attacker [165].
4) *Visual deception:* This PT, extracted from [166], repurposes benign visual elements to induce TRUST [167]. It leverages the OVERCONFIDENCE, TRUST, and HABITUATION PFs. Typosquatting and clone-phishing attacks exploit it by creating URLs that are visually similar to benign URLs.
5) *Incentive and motivator:* This PT, extracted from [168], encourages a desired behavior or compliance with a request. Incentive provides external rewards for action, while motivator provides internal rewards (i.e., gratification) for an individual. Incentive often leverages visceral triggers, which are commonly used in malvertising and click-baiting attacks, as well as in the Nigerian scam [169]. Motivator exploits SYMPATHY, EMPATHY, LONELINESS, and DISOBEDIENCE. Wire transfer scams exploit victims' SYMPATHY for the attacker as a motivator to encourage someone to transfer money to an attacker who claims to have made an erroneous money transfer.
6) *Urgency:* This PT, extracted from [167], refers to a situation which requires immediate action or is

ostensibly under time pressure [81], causing a decrease of chance in recognizing an attack [167]. It exploits the COGNITIVE MISER, FEAR, and NEGLIGENCE PFs. It is often used by scareware attacks to urge users to install software that detects threats (e.g., malware) or a plug-in that allows the user to view some desired contents [170].

7) *Attention grabbing:* This PT, extracted from [168], uses visual and auditory elements to prompt a victim to focus attention on deceptive attack elements to increase compliance. It exploits the ABSENTMINDEDNESS and CURIOSITY PFs. The malvertising, scareware, and click-baiting attacks exploit this PT along with visceral triggers and incentives to encourage compliance [170].
8) *Personalization:* This PT, extracted from [44], uses personal information to tailor messages or express similar interest to the victim to engender trust [171], [172]. It exploits the PERSONALITY and INDIVIDUAL DIFFERENCES PFs to increase the chance of success.
9) *Contextualization:* This PT, extracted from [44], projects an attacker as a member of the victim's group in order to establish commonality with the victim and increase the chance of attack success [53], [74]. This PT exploits the HOPELESSNESS, PERCEPTUAL CONTRAST, VULNERABILITY, IMPULSIVITY, CURIOSITY, and AFFECTIVE COMMITMENT PFs. It is often used in attacks like whaling, catfishing, and drive-by downloads [162].
10) *Quid pro quo:* This PT, extracted from [76], means "something for something else." It attempts to make a victim willing to take risks in exchange for a high payoff (e.g., money, free services, or avoiding embarrassment). It exploits the RECIPROCATION, GREED, and HOPELESSNESS PFs [55]. For example, the attacker can impersonate a police officer to make a victim pay for illegal content (e.g., pornography) on the victim's computer [38]; otherwise, the attacker threatens to arrest the victim for the possession of illegal content. In the Nigerian Prince Scam (419) [169], the PT incurs the expectation that one gives a small amount of money to receive a larger amount of money later.
11) *Foot-in-the-door*. This PT, extracted from [99], attains compliance for a large request by making small requests over time [173]. It exploits the CONSISTENCY PF. It is often used in the honey trap and catfishing attacks.
12) *Trusted relationship:* This PT, extracted from [174], exploits an existing trust relationship by taking advantage of the AUTHORITY, RESPECT, and TRUST PFs. For example, an attacker posing as a recruiter on LinkedIn, which is deemed by some as a trusted service provider, can connect to employment-seeking victims [175]; spamdexing (SEO) exploits a user's trust in a search engine provider (e.g., Google); and BEC exploits the trusted relationship between an executive and a subordinate employee.
13) *Affection trust:* This PT, extracted from [168], establishes an affectionate relationship with a victim. It exploits the AFFECTIVE COMMITMENT PF because affection makes an individual more willing to take risks and, thus, increases compliance even if it does not lower risk perceptions or increase trust [176]. It is commonly used in the catfishing and honey trap attacks.
14) *Decoy effect:* This PT, extracted from [177], attempts to make users believe that they are receiving a good deal (e.g., a deal with a lower-than-market price for some goods but never delivers when the victim pays upfront) [177], [178]. This PT exploits the TRUST, SCARCITY, AFFECTIVE COMMITMENT, and IMPULSIVITY PFs.
15) *Priming:* This PT, extracted from [179], attempts to influence an individual's subsequent decision through gradual manipulation (e.g., the attacker keeps sending a victim information about cryptocurrency being the next big thing before sending the victim a fake link to purchase cryptocurrency) [180]. This PT exploits the CURIOSITY, OPENNESS, TRUST, and GREED PFs.
16) *Loss aversion:* This PT, extracted from [102], is used when the attacker gives a victim something for free but then charges the victim enormously when the victim becomes attached to it (e.g., attacker tells a victim that the attacker has fake dollar bills that look real, then gives the victim real dollar bills as fake, and, finally, gives fake dollar bills in the final deal) [145]. This PT exploits the HABITUATION, SCARCITY, SOCIAL PROOF, CONSISTENCY AND COMMITMENT, PERCEPTUAL CONTRAST, FREEWHEELING, TRUST, GREED, CURIOSITY, and OPENNESS PFs.

Real-World Example: Fig. 6 describes a real-world email showing how an attacker exploits PFs via PTs, where purple highlights the AUTHORITY PF that is exploited by the persuasion PT, green highlights the impersonation PT, blue highlights the incentive & motivator PT, and yellow highlights the urgency PT. While we can determine the AUTHORITY PF via the persuasion PT, the other PTs can exploit multiple PFs out of the 46 PFs mentioned above. For example, the US$5.5 million incentive (i.e., the incentive & motivator PT) exploits the GREED and CURIOSITY PFs; the urgency PT exploits the IMPULSIVITY and ABSENTMINDEDNESS PFs; and the impersonation PT exploits the COGNITIVE MISER and ABSENTMINDEDNESS PFs.

As another example, we show that multiple PTs can be used together in a single SE attack. This is the Google and Facebook spear phishing scam [85], where the attacker creates a fake computer manufacturing company while pretending to work with Google and Facebook. The attacker sends spear phishing emails to targeted Google and Facebook employees, directing them to deposit money in the attacker's account for goods and services.


From: **Merrick Garland** <rosemarycormier14@gmail.com>
Date: Tue, Nov 23, 2021 at 9:57 AM
Subject: The United States Department Of Justice.
To:

Office name: Attorney Merrick Garland
Address: 66 John St 2nd floor, New York,
NY 10038, United States

Dear Beneficiary,

I am Attorney Merrick Garland, from the office of the attorney general of the United States. I am mandated to help you in release and processing of your payment. This is to inform you about your abandon fund $5.5 Million brought to our office on 13/10/2021 for confiscation due to your inability to claim your fund as the beneficiary.

We received many letters and emails from different people claiming that you sent them to receive this fund on your behalf, please confirm if you sent those people to receive your fund on your behalf.

Again, you have until 29/11/2021 to reclaim this fund If you still want to claim your fund Reply back on this email {merrickgarland377@gmail.com} and tell us the way you wish to receive your fund.

Your urgent feedback is needed.

Best Regards,

Attorney Merrick Garland,
From the Attorney General Office.
merrickgarland377@gmail.com

Fig. 6. ***Annotated real-world email showing how an attacker uses PTs to exploit PFs.***

From 2013 to 2015, the attacker collected a total of US$100 million. This attack exploits the impersonation PT by pretending to be a legitimate entity and the priming PT by gradually manipulating victims over time to believe that they are legitimate, while the attacker indeed addresses customer concerns as a legitimate company does.

Yet another example of SE attacks exploiting multiple PTs is the case of Hillary Clinton's campaign chairman John Podesta in the 2016 US presidential election. The attack is carried out by the attacker known as Fancy Bear, which is thought to be a Russian hacking group. The title of the email sent by the attacker is "Someone has your password" and the body of the email contains "You should change your password immediately" and "CHANGE PASSWORD" with a shortened `bit.ly` URL to click on. It also contains a timestamp and location stamp of Ukraine. The Clinton campaign professional thinks that the email is legitimate and asks Podesta to change his password [181]. He clicks on the URL to change his password, and as a result, the attacker logs into his email account to exfiltrate his emails. The attack exploits the visual deception and personalization PTs.

What Insight Can We Draw? The preceding systematization shows that the persuasion PT is the most studied. The six Principles of Persuasion have been widely adopted for business purposes (e.g., marketing), suggesting that attackers have been exploiting advanced knowledge for their malicious purposes. Moreover, one study [42] based on 200 malicious emails shows that the attention grabbing PT is most widely exploited in malicious emails.

Insight 6: Among the 16 PTs, attention grabbing is perhaps the most exploited, but persuasion is the most studied. Future studies may need to focus on the attention grabbing PT.

VII. SYSTEMATIZING ATTACKS

We systematize SE attacks based on their *objectives* (i.e., what an attacker attempts to accomplish) and *types* (i.e., how an attacker accomplishes its objectives). Note that one attack may have multiple objectives.

A. Attack Objectives

Our analysis of the references prompts us to categorize SE attacks based on three intents (or motives): *money*, *data*, and *recognition*, which can be, respectively, divided into two, five, and two objectives (i.e., nine objectives in total).

For attacks motivated by money, there are two objectives.

1) *Stealing money*, namely, an attacker attempts to steal money from victims. For example, BEC is often used to steal or extort money from victims [182]. An APWG report [3] shows that 27.7% of phishing attacks in the fourth quarter of 2022 target financial institutions with the intent to steal their money, and there is a 6% increase in attacks against payment processors such as PayPal, Venmo, and the VISA card company.
2) *Blackmailing*, namely, an attacker intends to obtain damaging information on its enemies or rivals to force them to do something for the attacker or get an upper hand on their enemy or rival [183]. For example, a Pakistani cybercrime group uses victims' personal WhatsApp data to blackmail them for money [183].

For attacks motivated by data (i.e., secret data), there are five attack objectives.

1) *Getting access to secure systems*, namely, an attacker uses SE attacks as a first step of full-fledged attacks for penetrating into secure networked systems to exfiltrate secret data (e.g., advanced persistent threats [184]).
2) *Stealing sensitive information*, namely, an attacker attempts to steal sensitive information from a user. When the user is a company or enterprise, the sensitive information can be the information of their customers (but not their trade secrets that will be designated as a different objective). For example, phishing is often used to steal sensitive information, such as passwords [185].
3) *Industrial espionage*, namely, an attacker attempts to spy on companies, enterprises, organizations, or individuals, and then compile the information available to the attacker, rather than stealing trade secrets [100].
4) *Stealing trade secrets*, namely, an attacker attempts to steal trade secrets of companies and enterprises for its own use or sell them to others [186].

5) *Game! Fun! Hobby!*, namely, an attacker attempts to lure an innocent, but often capable, person to do something for the sake of having fun or taking on a challenge for breaking into a system to steal and post information to prove their capabilities, while the attacker collects and abuses the information [187].

For attacks motivated by recognition (i.e., social recognition), there are two objectives.

1) *Fame and notoriety*, namely, an attacker attempts to become respected and looked upon by others as cybernerds, by carrying out SE attacks to prove their skills to their peers for recognition [40]. Unlike the objective of *Game! Fun! Hobby!* where a capable individual is actually exploited by the attacker to wage attacks, the present objective is about a capable individual showing off their skillset.
2) *Revenging*, namely, an attacker attempts to take revenge against enterprises, organizations, or individuals by releasing damaging information about them [187].

What Insight Can We Draw? It would be interesting to know what attack motive is the most popular and which sector is most targeted by SE attacks. First, the most popular attack motive appears to be money, as evidenced by the fact that attackers often target financial institutions. Since these institutions have taken tremendous steps to harden their cybersecurity, attackers appear to turn their efforts against nonfinancial institutions, which often have weak cybersecurity measures [86]. For example, a 2022 APWG report [3] shows that when attackers have individuals' personal identification information, they often exploit such information to receive gift cards from victims, with 60% requesting Amazon gift cards. Second, the financial sector is the primary targeted sector. In addition, attackers have turned to target the healthcare sector, primarily for the purpose of garnering information about individuals [15], [52].

Insight 7: Money is the most popular attack objective, and the finance sector is most targeted by SE attacks.

B. Attacks Types

As highlighted in Fig. 7, we divide SE attacks based on the medium they leverage into three categories: email versus website versus OSN, which, respectively, have sux, 12, and eight attack types (i.e., 26 attack types in total).

1) Email-Based Attacks: This category includes sux attacks, which are varying flavors of phishing.

1) *Generic phishing:* This attack sends phishing emails without a particular target in mind while hoping that some individuals will fall victim to them (i.e., no personalization in such phishing emails) [35], [188], [189]. A phishing email contains bait and hopes that someone will go for the bait and often intends to steal money. Based on our analysis of [190], this attack may use the incentive & motivator, urgency, and impersonation PTs to exploit the GREED PF because it attempts to entice victims for rewards such as the 419 scam that promises a large amount of money if a victim pays a small amount of money. This attack may also leverage the pretexting PT to present a scenario, whereby the victim agrees with the attacker and follows with the demand of the attacker. Phishing may continue to grow because of the limited success of defenses [35].
2) *Spear phishing:* A spear phishing email contains information personalized for a specific target, usually addressing the target by name and title. It often intends to steal money, get access to systems, steal sensitive information, or revenge on an institution or individual. This attack may use the personalization and impersonation PTs to exploit the AUTHORITY PF to deceive a recipient into believing that the attacker is an authoritative figure, and thus, the recipient must act promptly [38], [162], [191]. Attackers are willing to make a big effort to wage this attack. For example, a study [35] based on 370 million spear phishing emails shows that over 60% of source addresses send three or fewer spear phishing emails, and over 40% of source addresses send exactly one spear phishing email (i.e., most attackers hardly reuse any email addresses to send spear phishing emails, and even if they do, not more than three times).
3) *Clone phishing:* A clone phishing email is cloned from a previously sent/received email, by replacing its links and/or attachments with malicious ones and spoofing the legitimate sender's email address so that the target would not suspect the legitimacy of the email [192], [193], [194]. This attack often intends to steal money and sensitive information. Based on our analysis, this attack may use the impersonation and visual deception PTs to exploit the TRUST PF because the attacker attempts to make a victim think that a cloned email is a continuation of a previous communication [192], [193].
4) *Whaling:* A whaling email is similar to a spear phishing email by targeting specific individuals. Unlike

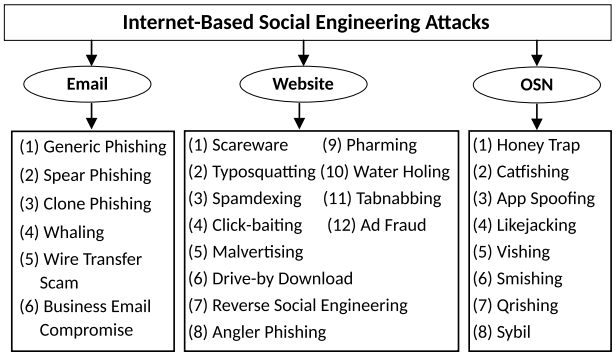


Fig. 7. ***Attack types based on the medium that is leveraged to wage an attack: email versus website versus OSN.***

spear phishing that can target any individuals, whaling emails target management, such as CEOs [5], [38], [162]. It often intends to steal money, get access to systems, steal sensitive information, or for revenge. Based on our analysis of [38], [162], this attack may use the personalization and impersonation PTs to exploit the TRUST PF because the attacker attempts to deceive, for example, a CEO into believing in the content of an email. According to Goel and Jain [162], 55% of the organizations in their dataset observe an increase in whaling attacks in 2016, and 13% of them indicate that these attacks have a very significant impact on their organization.

5) *Wire transfer scam:* This attack sends an email to targeted individuals to deceive them into sending money (via, for example, Western Union) to pay for services or goods [195]. An attacker often impersonates a service company, such as a utility, to threaten that a victim's services will be cut off immediately unless a wire transfer is made while sometimes impersonating reputable individuals [196]. It intends to steal money. Based on our analysis of [195], this attack may use the visual deception and impersonation PTs to exploit the FEAR PF because the attack threatens to cut victims' services. The ease of conducting a wire transfer scam is demonstrated in [195], which suggests that the most important factor in stopping the transfer of money is time (because money transfer is almost instantaneous).
6) *BEC:* This attack uses email frauds against private, government, and nonprofit organizations, by targeting specific employees with spoofed emails impersonating a senior colleague, such as the CEO or a trusted customer [197], [198]. This attack often intends to steal money [197]. Based on our analysis of [197], this attack may use the impersonation, urgency, visual deception, and personalization PTs to exploit the TRUST PF because the attacker attempts to deceive victims into believing that they are paying a legitimate bill for goods/services from a trusted party. This attack could cost victims a lot of money, as shown by the cases of Facebook and Google [197].

2) What Insight Can We Draw?: The preceding systematization suggests that email-based SE attacks have exploited 44 PFs through 11 PTs. APWG reports that email has been the most widely used SE attack over the past years [3]. FBI reports that BEC is the most damaging SE attack in terms of the financial loss that it incurs, as companies lost $12 billion to BEC from 2013 to 2018 [197]. APWG also reports that the average amount requested in wire transfer from a BEC attack in the fourth quarter of 2022 is $132,559, which is a 41% increase from the third quarter of the same year [3]. This leads to the following.

Insight 8: Email is widely used in SE attacks, among which BEC attacks cause the most significant financial loss.

3) Website-Based Attacks: This category includes 12 attacks. These attacks are not necessarily complementary or orthogonal to each other because one attack may leverage another as a supporting technique (e.g., *Ad Fraud* may use *malvertizing* as a support technique).

1) *Scareware:* This attack pops up a window with warning content, which tells the user that the computer has been infected by malware and the user should click a link or call a number shown on the pop-up window to get help [199]. The attacker intends to scare the user into clicking the link or calling the number so that the attacker can collect the user's sensitive information or ask the user to send a gift card number to the attacker to have the problem fixed remotely. Most scareware does not harm the computer in question [200]. This attack uses the attention grabbing, urgency, and persuasion PTs to exploit FEAR PF because the attacker scares victims into believing that their computer is compromised and needs immediate attention. According to Miramirkhani et al. [199], scammers use thousands of domains and phone numbers to scare victims to call them; scammers also use JavaScript techniques to make it harder for victims to navigate away from a scareware message; the five most frequently used words in scareware messages are Techsupport, Alert, PC, Security, and Windows.
2) *Typosquatting (or URL spoofing):* This attack takes a user to a malicious website when the user mistypes a URL, such as `www.bank0famerica.com` for `www.bankofamerica.com`, where the former mimics the latter while incorporating a malicious payload [38]. Based on our analysis of [38], this attack uses the visual deception and impersonation PTs to exploit the NEGLIGENCE PF because it anticipates individuals mistyping. This attack uses simple but effective cosmetic deceptions and can be achieved by registering domain names that are similar to popular, legitimate websites and are possibly misspelled by users, such as the domains mentioned above.
3) *Spamdexing (or* search engine poisoning): This attack tricks a search engine into making a malicious website on the top of the list returned by a search [38]. It is often exploited by technical support scammers [201], and it is effective because many users trust the search results listed on the top and treat them as most relevant, causing them to most likely visit these websites. One example of spamdexing is searching for the "best free video recorder app" where the attacker crafts a website with the top ten free video recorders and links to download them, but the links can be a source of drive-by download where malware is downloaded to a victim's computer. Based on our analysis of [38], this attack uses the impersonation and visual deception PTs to exploit the TRUST PF because the attacker anticipates that users treat the websites on the top of the list of search results as most relevant. Our analysis of [201] also indicates that this attack uses the

urgency PT (via continuous pop-up messages/dialog) to exploit: 1) the FEAR PF as the user panics and follows the attacker's recommendation and 2) the NEGLIGENCE and IMPULSIVITY PFs when the victim does not use due diligence.

4) *Click-baiting:* This attack places an enticing text/image on a web page to draw the attention of visitors so that they click on a link to a malicious or compromised website [202]. One example is a message on a website reading "Betty reveals how she gets to 100 years of age without ever doing sports." When users click on the link, it takes them to the website with a made-up story about Betty's secret of old age, but the website actually hosts malware to infect victims' computers. Note that click-baiting is not necessarily spam or fake because news outlets may use this technique [203]. This attack is often motivated by the attacker's desire to increase the traffic to their websites where users fall victim according to the objective of the attacker (e.g., download malware or sell bogus goods). Based on our analysis of [203], this attack can use the attention grabbing, visual deception, and incentive & motivator PTs to exploit the CURIOSITY PF because it entices victims to click on the link to figure out more information.
5) *Malvertising:* It is a major culprit that exposes users to technical support scams [199]. It abuses advertisements to spread malware such that when a user clicks on the advertisement, the user may be redirected to a malicious website [32], [204]. This can also be achieved by pushing ad notifications to users [205]. This attack is motivated to conceal the real intent behind an ad. Based on our analysis of [199], this attack uses the attention grabbing, visual deception, and urgency PTs to exploit: 1) the TRUST PF because the attacker attempts to make the victims believe that they are getting legitimate ads; 2) the NEGLIGENCE PF when the victims do not practice due diligence; and 3) the CURIOSITY PF when victims want to see the details of the ad. Moreover, the attacker may further use the foot-in-the-door PT to exploit the CONSISTENCY PF [199].
6) *Drive-by download:* This attack compromises a browser when one visits a malicious or compromised website, possibly prompted by a phishing email containing the malicious URL [206]. It intends to make victims download malware to their computers. Based on our analysis of [32], this attack can use the trusted relationship and urgency PTs to exploit: 1) the TRUST PF because a victim may trust the website in question; 2) the VULNERABILITY PF when a victim is not aware of this attack; and 3) the NEGLIGENCE PF when a victim does not update/patch a browser or does not pay attention to recognizing malicious websites.
7) *Reverse SE:* This attack attempts to gain the trust of victims before executing the attack. This attack tricks a user into contacting the attacker who then uses the opportunity to pursue their motive. More specifically, it is called reverse SE because it creates a situation that causes a victim to contact the attacker [207], for example, by providing a free online streaming service to prompt users to contact the attacker for the service [208]. This attack intends to have victims contact the attacker. Based on our analysis of [207], this attack uses the trusted relationship and urgency PTs to exploit the TRUST PF because it puts a victim in a situation of need, thereby causing them to contact the attacker. This attack is unique in that it creates trust between a victim and makes the victim believe that the attacker is a legitimate entity for a legitimate purpose, and, as a result, the victim contacts the attacker and gets exploited later [207].
8) *Angler phishing:* This attack attempts to lurk comments posted by users on social forums such as yelp and then takes advantage of any comment that may need a resolution [77]. For example, an attacker may see a comment from a customer complaining about a purchase. The attacker then poses as a customer satisfaction specialist and asks the customer for detailed information in order to address the complaint [209]. An unsuspecting customer may give away personal information in hoping of a resolution [9], [210]. This attack intends to extort brand customers, especially disgruntled customers who go online to air their grievances. Based on our analysis of [210], this attack uses the trusted relationship and contextualization PTs to exploit: 1) the VULNERABILITY PF because frustrated victims desperately need solutions and 2) the TRUST PF because victims put trust in the service company in question. To see how significant the attack is, we mentioned that a survey shows that some well-known brands, such as Amazon, Nike, and Samsung, suffered from an 1100% increase in this attack from 2014 to 2016, where attackers pose as legitimate brand customer representatives to take advantage of unsuspecting customers [210]. This attack may leverage the priming PT to gradually prepare the individual in the interaction with the attacker until the attacker gets an opportunity to strike.
9) *Pharming:* This attack builds malicious websites to steal money or sensitive information from victims when they visit these websites [114]. This attack is motivated to steal personal and financial information from its victims. Based on our analysis of [114], this attack uses the impersonation and visual deception PTs to exploit: 1) the TRUST PF because victims do not think that these websites are malicious and 2) the NEGLIGENCE PF because victims do not perform due diligence. This attack is unique in the sense that it injects fake information into the domain name system (DNS) server of the website that may be visited by victims [211].

10) *Water holing:* This attack exploits vulnerabilities of third-party websites to attack victims when visiting them [99]. For example, an attacker may want to attack a company but cannot break the network security to penetrate the company's network. Because of this, the attacker can compromise a website that employees of the target company regularly visit and set a trap (water hole). When an employee of a targeted organization visits a compromised website, which is considered safe, and their link is clicked, the employee's computer is compromised and then leveraged to attack others. This attack often intends to steal money or sensitive information. Based on our analysis of [29], this attack uses the trusted relationship PT to exploit the TRUST PF because victims believe that the websites they are visiting are secure.
11) *Tabnabbing:* This attack attempts to deceive a victim into visiting a malicious website, which mimics a legitimate website and asks the victim to login into the malicious website, while making the victim think that the malicious website is legitimate and forwarding the victim's login credentials to the legitimate website [5]. It often leverages the same origin policy of browsers, where a second page on a browser can access scripts from another page as long as both pages have the same origin [212]. It intends to steal sensitive information (e.g., login credentials). Based on our analysis of [5], this attack uses the trusted relationship and visual deception PTs to exploit: 1) the ABSENTMINDEDNESS PF because the attack makes a victim believe that a previously visited website is asking for login credentials again and 2) the TRUST PF because the attack makes a victim believe that the victim needs to reenter the tab that is left open.
12) *Ad fraud:* This attack defrauds advertisement, where the fraudster deceives a victim to advertise its services by generating fake traffic (possibly via malvertising, scareware, click-baiting, and likejacking) [213]. It is often motivated to steal money in the sense that the advertisement does not incur real traffic from real users but forged traffic instead. Based on our analysis of [213], this attack uses the contextualization and impersonation PTs to exploit the TRUST PF because the attack makes victims believe that they are getting legitimate traffic to their advertisements.

What Insight Can We Draw? The preceding systematization shows that website-based SE attacks have exploited 38 PFs through 13 PTs. However, we do not observe the exploitation of the personalization PT by website-based SE attacks, perhaps because: 1) unlike emails, it may be difficult to develop a malicious website specific to an individual and 2) the mere fact that a website is personalized to a single individual may be counterintuitive, as it may instead raise the suspicious of the individual. On the other hand, the impersonation PT appears to be the main driver behind website-based SE attacks, perhaps because users used to deal with legitimate entities (i.e., whom the attacker claims to be).

Insight 9: Among the PTs exploited by website-based SE attacks, impersonation is the most widely exploited one.

4) OSN-Based Attacks: This category includes eight attacks.

1) *Honey trap:* This attack targets a particular victim with a love-related relationship and may be seen as the counterpart of spear phishing [76], [214], [215]. For example, John knows that Philip likes blonds and, thus, creates a fake profile of a blond on Instagram to like and comment on Philip's posts; Philip sees a blond liking his posts and thinks that it is an opportunity for him to meet a blond; once a relationship is established, John can deceive Philip in many ways, including financial extortion [216]. This attack is motivated to lure victims into a romantic relationship for later extortion. Based on our analysis of [76], this attack uses the impersonation, affection trust, and persuasion PTs to exploit the LONELINESS PF because lonely people often seek attention on electronic platforms.
2) *Catfishing:* This attack creates a fake persona to seek online dating to lure victims interested in the persona, similar to generic phishing because the attack does not target a specific victim [217]. For example, the attacker posts as woman to lure men to send them money for made-up reasons, for example, "My Internet service will be suspended for accumulated bills, please help me pay or I'll not be able to chat with you if my Internet is suspended." This attack intends to extort money from victims. Based on our analysis of [218], this attack uses the impersonation and persuasion PTs to exploit: 1) the LIKING (SIMILARITY) PF because victims have the tendency to react positively to someone with whom they have a relationship [95] and 2) the LONELINESS PF because lonely people tend to seek online friendship [152]. It is interesting to note that catfishing became popular in 2010 after the movie Catfish, and 38% of men studied tend to catfish [218]. Note that a major difference between the honey trap attacker and the catfishing attacker is that the former poses as a very attractive person or a celebrity, but the later pretends to be an average individual looking for romance [76].
3) *App spoofing:* This attack uses bogus apps to spoof legitimate ones on less-regulated platforms. When a user uses the same credential for multiple platforms, the attacker can steal a user's credentials to get access to the user's account on other platforms [219]. This attack is usually carried out on social media where malicious or compromised users send links to their followers to encourage them to download spoofed apps. The attacker can also post the link to download the app in comment sections of online social media

platforms. This attack is often motivated to spread spoofed apps to collect victims' login information for the legitimate apps that are being spoofed. Based on our analysis of [220], this attack uses the visual deception and impersonation PTs to exploit the OPENNESS and CURIOSITY PFs as users who are open and curious often try new things.

4) *Likejacking:* This is the social media version of a clickjacking attack. This attack places a transparent layer (e.g., transparent iframe) on a legitimate webpage so that when a user clicks anywhere on the webpage, the user is actually clicking on the transparent layer, which directs the user to the attacker's website [20], [221].
   In likejacking, when a user sees the "like" button on a Facebook post, on top of which there is a transparent layer not visible to the user, the user may click on the page and then be directed to a malicious website. Based on our analysis of [20], this attack uses the visual deception, attention grabbing, and persuasion PTs to exploit: 1) the LIKING AND SIMILARITY PF because the attacker sets the trap knowing that people tend to like comments of people that they follow on OSN and 2) the IMPULSIVITY PF because users tend to click on everything that they like on social media.
5) *Vishing:* This attack is the use of voice over IP (VoIP) calls to impersonate a legitimate entity. For example, an attacker can use vishing together with PFs such as AUTHORITY and SCARCITY to make its victim believe in the attacker [101]. The voice content attempts to trick a victim into performing harmful actions to benefit the attacker [94]. For example, an attacker may pretend to be a student loan service provider, present a loan repayment deal, and ask a victim for personal information, such as a bank account and credit card number. An attacker may also use Caller-ID-Spoofing to spoof a legitimate phone number to trick the receiver into believing that the phone call is coming from a legitimate person or company. This attack is often motivated to get immediate access to a victim's login information. Based on our analysis of [101], this attack uses the impersonation and persuasion PTs to exploit: 1) the TRUST PF because victims often believe the callers who claim they are and 2) the AGREEABLENESS PF because victims are often kind by nature and easily trust people. It is interesting to note that lightweight authority figures (e.g., bankers) may have a lesser chance of making victims give away their sensitive information to callers but heavyweight authority figures (e.g., police officers) would have a higher chance of making victims give away their sensitive information to callers [101].
6) *Smishing:* This attack uses mobile apps to send impersonating messages to lure victims into divulging sensitive information to the attacker [20], [222]. For example, an attacker may send a text message to pretend to be a bank and ask a victim to update their PIN while possibly using Caller-ID-Spoofing to make the attack hard to recognize by a victim [20]. Even people who do not answer unsolicited messages may react to smishing messages because they look like texts from real persons. This attack often intends to trick people into downloading malware into their mobile devices or giving their sensitive information to the attacker. Based on our analysis of [20], this attack uses the personalization, impersonation, incentive & motivator, and urgency PTs to exploit: 1) the AUTHORITY and TRUST PF because victims would trust that they are dealing with a person of authority; 2) the GREED PF because victims often want something for free; and 3) the IMPULSIVITY PF because victims may quickly react to an attack message without a second thought. This attack may also leverage: 1) the decoy effect PT to prepare the victim while waiting for the right opportunity to strike and 2) the loss aversion PT to provide the victim with correct information over time until the attacker strikes.
7) *QRishing:* This attack exploits QR codes to deceive victims to visit bogus websites, which may mimic legitimate websites to collect sensitive information (e.g., login information) [20] or even spread malware [29], [52]. It can be disseminated in many ways, such as sending phishing QR codes in emails, posting QR codes on OSN platforms or websites, and disseminating QR codes via hard paper copies. This attack often intends to provide links for victims to download malware into their mobile devices. Based on our analysis of [20], this attack uses the visual deception and persuasion PTs to exploit the TRUST, SOCIAL PROOF, and IMPULSIVITY PFs. Note that QRishing is popular because, nowadays, most people use mobile devices to do business [20].
8) *Sybil attack:* The attacker attempts to clone or mimic legit users to dupe their followers on social media (e.g., Facebook, Twitter, and Instagram) [111]. Moreover, the attacker may create multiple fake identities on a single OSN platform, also known as social Bots [174], [223]. This attack is often motivated to gain a high social media presence before launching another attack (e.g., catfishing). Based on our analysis of [111], this attack uses the impersonation, trusted relationship, and contextualization PTs to exploit: 1) the TRUST PF because humans trust their friends; 2) the SOCIAL PROOF PF because humans often behave as their friends do; and 3) the IMPULSIVITY PF because humans are often interested in making new friends.

What Insight Can We Draw? First, the preceding systematization suggests that OSN-based SE attacks have exploited 41 PFs via 12 PTs. The three types of attacks, or 26 attacks in total, collectively exploit the 46 PFs through

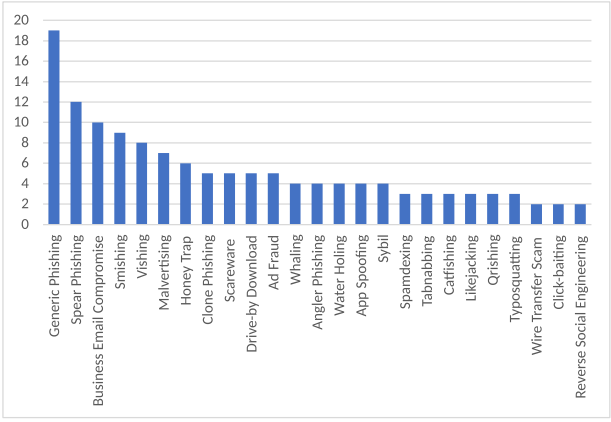


Fig. 8. ***Number of references per attack studied in multiple papers (e.g., generic phishing is studied in 19 papers).***

the 16 PTs. This means that SE attackers are extremely proactive in exploiting PFs.

Second, we want to see which attacks have been studied most. Fig. 8 plots the number of references on attacks. Among the 135 papers studying attacks, 49 (or 37.8%) investigate various flavors of phishing, with 19 papers on generic phishing, 12 on spear phishing, ten on BEC, five on clone phishing, four on whaling, and two on wire transfer scam. This suggests that phishing has been studied most, perhaps because it accounts for 90% of data breaches [7]. This means that there is an inherent skewness in the literature for our study.

Insight 10: Attackers are extremely proactive in exploiting PTs and PFs. Phishing is the most widely investigated attack but remains the most damaging attack, while many attacks are much less studied.

VIII. SYSTEMATIZING DEFENSES

We divide defenses into three categories based on the attacks that they are defending against: email-based versus website-based versus OSN-based. It may be tempting to present defenses against each attack mentioned above, but it is less effective because one defense may be applicable to multiple attacks. Since our study is based on the psychological lens of PFs and PTs, we focus on the defenses that leverage PFs and/or PTs. There are 12 defenses that consider PTs and/or PFs, including four against email-based SE attacks, five against website-based SE attacks, and three against OSN-based SE attacks.

A. Defenses Against Email-Based Attacks

Based on the references systematized, there are only four defenses that consider PFs and PTs. The first defense is *BEC Guard* [197]. This defense is a machine learning-based method for detecting BEC emails by leveraging email content-based features that are related to: 1) the urgency PT because it uses urgency cues in email content (e.g., "*Joe, I need your urgent help?*), 2) the impersonation PT because it copes with impersonation in emails; and 3) the personalization PT because it explicitly detects personal identification information. This defense leverages: 1) the COGNITIVE MISER PF because this PF is exploited by the urgency PT to trigger people to use mental shortcuts in reasoning; 2) the AUTHORITY PF because this PF is exploited to craft BEC emails; 3) the RESPECT PF because people often respect authoritative figures; and 4) the TRUST PF because people often trust authoritative figures. This defense achieves a precision of 98.2% and a false-positive rate of less than one in five million, highlighting the effectiveness of PF- and PT-based defenses.

The second defense is *PhishNet-NLP* [224]. This defense uses natural language processing techniques to detect phishing emails while defining email content-based features to capture cues related to the urgency, incentive & motivator, urgency, and incentive & motivator PTs, such as sentences that create a sense of urgency, worry, threat, and concern or sentences that offer an incentive to users. This defense also leverages: 1) the COGNITIVE MISER PF to prevent individuals from using mental shortcuts in reasoning; 2) the FEAR PF to prevent people from fearing missing a deadline; 3) the NEGLIGENCE PF to prevent people from neglecting due diligence; and 4) the CURIOSITY PF to prevent humans from falling victim to malicious incentives. This defense achieves a 97% detection rate and a very low false-positive rate.

The third defense is *L-XGB* [225]. This defense uses the long short-term memory (LSTM) model and the extreme gradient boosting tree (XGBoost) technique to detect phishing emails while leveraging email content-based features that are related to: 1) the attention grabbing PT because the features correspond to cues of the attention grabbing PT, such as the variation of the fonts in a malicious email and 2) the incentive & motivator PT because the subject line in malicious emails usually carries an incentive or a hint of the incentive upfront. Our examination further shows that the defense leverages: 1) the ABSENTMINDEDNESS PF because it prevents emails with enticing subject lines from reaching their recipients; 2) the GREED PF that is exploited by the incentive and motivator PT; 3) the RECIPROCATION PF because people use paying-back as a motivator to comply with a request of goodwill; and 4) the CURIOSITY PF because humans are often curious when presented with an incentive. This defense achieves a 98.58% precision in detecting phishing emails.

The fourth defense is *KM-SMOTE* [226] against spear phishing. It uses email content-based features that are related to the impersonation PT to deal with attackers impersonating a known person. Our examination of the defense shows that it also leverages the AUTHORITY, RESPECT, and TRUST PFs to mitigate attackers posing as authoritative figures. It achieves a 93.55% precision in detecting spear phishing.

What Insight Can We Draw? There are many defenses against email-based SE attacks that do not consider

psychological attributes (e.g., [188], [191], [227], and [228]). However, there are only four defenses that leverage ten PFs and four PTs in total, in sharp contrast to the 44 PFs and 11 PTs that have been exploited by email-based SE attacks. This could explain why existing defenses are not effective enough, extending what is observed in [175], namely, that existing defenses against spear phishing are ineffective because they fail to account for the human cognitive factors exploited by attacks.

Insight 11: There is a discrepancy between email-based SE attacks and defense efforts at using PFs and PTs.

B. Defenses Against Website-Based Attacks

Based on our systematization of the references, there are five website-based defenses that consider PFs and PTs. The first defense is *Phishpedia* [229]. This defense uses deep learning to detect phishing websites while comparing their logos and variant logos. Our examination of the defense shows that it leverages: 1) the visual deception PT because it considers cues related to the visual deception PT, such as comparing the presented logos to the legitimate ones and 2) the impersonation PT because it detects fake websites impersonating legitimate ones via feature matching between the fake websites and the legitimate websites. Our examination of the defense shows that it also leverages: 1) the ABSENTMINDEDNESS PF because it eliminates fake websites impersonating legitimate ones, thereby preventing users from absentmindedly visiting malicious websites; 2) the AUTHORITY, RESPECT, and TRUST PFs because it mitigates people's blind belief, respect, and trust in authoritative figures; and 3) the HABITUATION PF because people who have the habit of visiting a website may not detect a fake one of it. This defense achieves a 98.2% precision in detecting phishing webpages.

The second defense is *VisualPhishNet* [230]. This defense uses deep learning to detect phishing websites. It leverages: 1) the visual deception PT by considering the visual similarities between the legitimate websites and the fake websites and 2) the impersonation PT by detecting impersonating websites that are not on the trusted list. It also leverages: 1) the OVERCONFIDENCE PF because it mitigates people's confidence in a website that they frequently visit that could make them neglect suspicious clues exhibited by a malicious website; 2) the AUTHORITY, RESPECT, and TRUST PFs because it mitigates people's blind belief, respect, and trust in authoritative figures; and 3) the HABITUATION PF because people frequently visiting a website may not notice that they are visiting a fake one that mimics the real one. This defense achieves a 98.79% precision rate in detecting zero-day phishing.

The third defense is *PROTECT* [231]. It uses a game known as PERSUADED [232] to train employees against persuasion at the work place. The game exposes a trainee to challenging situations to build secure responses to attacks. It leverages: 1) the persuasion PT as indicated by the game name and 2) the impersonation PT because it uses both virtual and physical impersonation in the game. The defense further leverages: 1) the AUTHORITY PF because it has a training scenario where the attacker assumes authoritative figures; 2) the VIGILANCE PF because it is geared toward making employees vigilant against SE attacks; and 3) the SELF-EFFICACY PF because it trains employees to detect SE attacks. The study does not report any quantitative effectiveness.

The fourth defense is *BaitAlarm* [233]. This defense leverages that the visual appearance of a webpage is reflected by its page layout and contents by computing the similarity of two webpages via their layouts and contents to detect phishing websites. It leverages the visual deception PT because it uses visual similarity cues (e.g., logos). It also leverages: 1) the OVERCONFIDENCE PF because it deals with humans' confidence toward websites; 2) the TRUST PF because people trust entities that they often interact with but do not know that entities can be impersonated; and 3) the HABITUATION PF because victims often do not pay much attention to the websites that they frequently visit. This study reports a 100% detection rate and a 0% false-negative rate.

The fifth defense is *hybrid phishing detection* [234]. It applies machine learning to analyze webpage contents such as logos and integrates visual identity with textual identity to detect phishing websites. It leverages: 1) the visual deception PT because it considers visual similarity cues such as logos; 2) the attention grabbing PT because it uses textual cues such as font size and texture; and 3) the impersonation PT because it uses visual and textual cues to deal with the impersonation of legitimate brands or entities. It also leverages: 1) the OVERCONFIDENCE and HABITUATION PFs because it reduces the chance that people do not pay a due amount of attention in recognizing fake ones and 2) the TRUST and AUTHORITY PFs because it detects the presence of authoritative figures to mitigate their blind trust in such figures. This defense achieves a 98.6% accuracy in detecting phishing websites.

What Insight Can We Draw? There are many defenses against website-based attacks that do not consider psychological attributes (e.g., [235] and [236]). However, there are only five defenses that leverage nine PFs and three PTs in total, in sharp contrast to the 38 PFs and 12 PTs that have been exploited by website-based SE attacks in total.

Insight 12: There is a discrepancy between website-based SE attacks and defense efforts at using PFs and PTs.

C. Defenses Against OSN-Based Attacks

There are only three defenses (against OSN-based SE attacks) that consider PFs and PTs. The first defense is *OSN Profile Cloning Protection* [237]. It detects profile cloning attacks by leveraging similarities between OSN accounts

in terms of their attributes, friend lists, and strength of ties. It leverages the impersonation PT by using attribute similarity, friend list similarity, and strength of ties. It leverages: 1) the AUTHORITY PF by detecting intra-site cloning of authoritative figures and 2) the TRUST PF by analyzing the trust relationship with cloned profiles. The defense could be enhanced to prevent: 1) honey trap that involves a fake social media account impersonating a popular individual (e.g., celebrity) and 2) catfishing that often involves a fake social media account by detecting fake accounts. It achieves a 97.23% precision in detecting OSN-based profile cloning attacks.

The second defense is *CSE-PUC* [238]. This defense uses natural language processing techniques to detect persuasive words in chats. It leverages the persuasion PT because it detects persuasive cues in messages. It also leverages the six PFs corresponding to the six principles of persuasion by detecting the pertinent cues (e.g., "limited offer" pertinent to the SCARCITY PF and "like" or "admire" pertinent to the LIKING PF). It achieves a 71.63% precision in detecting OSN-based SE attacks, suggesting room for improvement.

The third defense is *SEADer++* [239]. It uses natural language processing to parse and check for grammatical errors in conversation text. It leverages the persuasion PT by detecting persuasive cues in messages. It also leverages: 1) the AUTHORITY PF because it detects message content pertinent to authority such as "Police," CEO, and names and seals of government agencies and 2) the SCARCITY PF because it detects cues related to the exploitation of a word or phrase depicting limited supply of something. It achieves a 92.6% precision in detecting OSN-based SE attacks.

What Insight Can We Draw? There are many defenses against OSN-based SE attacks (e.g., [111], [126], and [223]). However, there are only three defenses (against OSN-based attacks) that leverage nine PFs and three PTs in total. By contrast, OSN-based SE attacks have exploited 41 PFs and 12 PTs in total.

Insight 13: There is a discrepancy between OSN-based SE attacks and defense efforts at using PFs and PTs.

IX. SYSTEMATIZING RELATIONSHIPS BETWEEN PFs, PTs, ATTACKS, AND DEFENSES

To further systematize the discussion in the preceding sections, we map the relationships between the PFs, PTs, attacks, and defenses. The mapping presents a succinct representation of the state-of-the-art knowledge in this field. However, the lack of quantitative results in the literature (e.g., the impact of PFs) prevents us from conducting a quantitative meta-analysis. As highlighted in Fig. 9 and elaborated on in the following, the mapping describes: 1) which PTs exploit which PFs; 2) which attacks exploit which PTs; 3) which defenses leverage which PFs; and 4) which defenses address which attacks.

A. Which PTs Exploit Which PFs?

A PT can exploit multiple PFs, which may fall into one or multiple psychological categories. We discuss this according to the five categories of PFs.

First, the PTs that exploit *cognitive* PFs include: 1) urgency exploits the COGNITIVE MISER and EXPERTISE PFs; 2) attention grabbing exploits the ABSENTMINDEDNESS PF; and 3) visual deception exploits the OVERCONFIDENCE PF. That is, four (out of four) *cognitive PFs* have been exploited by PTs.

Second, the PTs that exploit *emotion* PFs include: 1) urgency exploits the FEAR and HOPELESSNESS PFs; 2) attention grabbing exploits the FEAR PF; 3) visual deception exploits the HOPELESSNESS PF; 4) incentive and motivator exploits the GREED, FEAR, SYMPATHY, EMPATHY, and LONELINESS PFs; 5) quid pro quo exploits the GREED PF; and 6) priming exploits the GREED PF. That is, six (out of six) *emotion PFs* have been exploited by PTs.

Third, the PTs that exploit *social* PFs include the following: 1) urgency exploits the AUTHORITY PF; 2) attention grabbing exploits the SCARCITY PF; 3) incentive and motivator exploits the DISOBEDIENCE PF; 4) persuasion exploits the AUTHORITY, RECIPROCATION, LIKING, SCARCITY, SOCIAL PROOF, and CONSISTENCY PFs; 5) quid pro quo exploits the RECIPROCATION and SCARCITY PFs; 6) foot-in-the-door exploits the CONSISTENCY PF; 7) trusted relationship exploits the AUTHORITY and RESPECT PFs; 8) impersonation exploits the AUTHORITY, SIMILARITY, and RESPECT PFs; 9) contextualization exploits the LIKING (SIMILARITY) and PERCEPTUAL CONTRAST PFs; 10) decoy effect exploits the SCARCITY PF; and 11) loss aversion exploits the PERCEPTUAL CONTRAST PF. That is, nine (out of nine) *social PFs* have been exploited by PTs.

Fourth, the PTs that exploit PID PFs include the following: 1) urgency exploits the SELF-CONTROL, IMPATIENCE, IMPULSIVITY, and OPENNESS PFs; 2) attention grabbing exploits the CURIOSITY PF; 3) visual deception exploits the TRUST and VIGILANCE PFs; 4) incentive and motivator exploits the FREEWHEELING PF; 5) quid pro quo exploits the IMPATIENCE PF; 6) foot-in-the-door exploits the VULNERABILITY PF; 7) trusted relationship exploits the TRUST PF; 8) impersonation exploits the TRUST PF; 9) contextualization exploits the OPENNESS PF; 10) pretexting exploits the TRUST PF; 11) personalization exploits the DISORGANIZED, FREEWHEELING, INDIVIDUAL INDIFFERENCE, NEGLIGENCE, TRUST, SELF-CONTROL, VULNERABILITY, IMPATIENCE, IMPULSIVITY, SUBMISSIVENESS, CURIOSITY, LAZINESS, VIGILANCE, OPENNESS, CONSCIENTIOUSNESS, EXTRAVERSION, AGREEABLENESS, and NEUROTICISM PFs; 12) decoy effect exploits the FREEWHEELING, TRUST, and IMPULSIVITY PFs; and 13) priming exploits the IMPULSIVITY, CURIOSITY, and OPENNESS PFs. That is, 18 (out of 18) PID PFs have been exploited by PTs.

Fifth, the PTs that exploit *workplace* PFs include the following: 1) urgency exploits the WORKLOAD, STRESS,

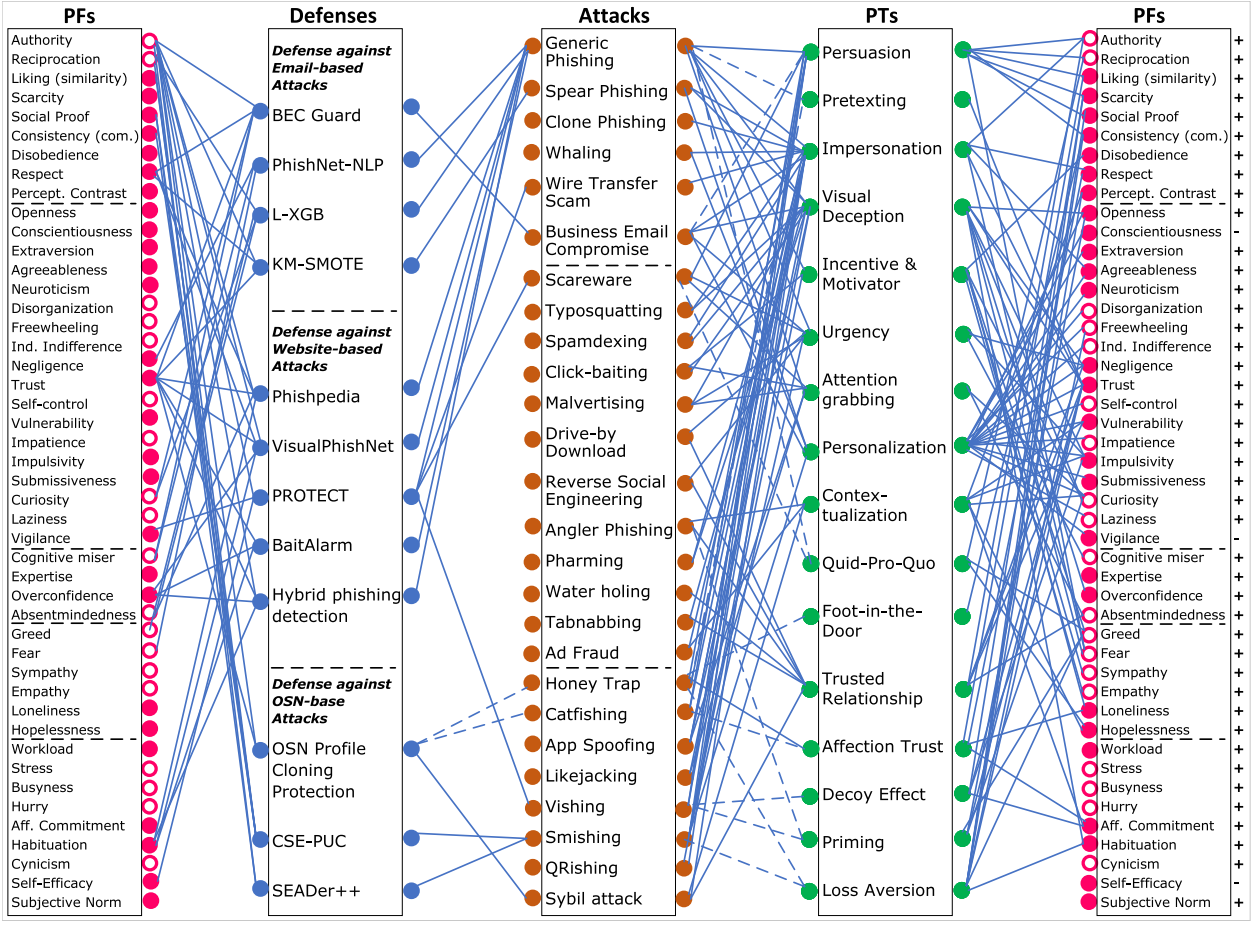


Fig. 9. ***Relationships between PFs, PTs, attacks, and defenses, where a dashed line inside a box indicates different categories, the PFs with empirical quantitative studies are represented with a filled circle and an empty circle otherwise, and the "+" ("−") sign indicates that a factor increases (decreases) human susceptibility to attacks. A solid line represents a relationship that is extracted from a reference and discussed in the text, and a dashed line represents a potential relationship that is not extracted from the references but based on our analysis. "Aff." is short for Affective, "Ind." is short for Individual, "com." is short for commitment, and "Percept." is short for Perceptual.***

and HURRY PFs; 2) visual deception exploits the HABITUATION PF; 3) trusted relationship exploits the HABITUATION PF; 4) impersonation exploits the WORKLOAD, AFFECTIVE COMMITMENT, and SUBJECTIVE NORM PFs; 5) contextualization exploits the STRESS PF; 6) pretexting exploits the HURRY PF; 7) personalization exploits the STRESS, HURRY, and SUBJECTIVE NORM PFs; 8) affection trust exploits the AFFECTIVE COMMITMENT PF; 9) priming exploits the CYNICISM and HABITUATION PFs; and 10) loss aversion exploits the STRESS, BUSYNESS, and HABITUATION PFs. That is, nine (out of nine) *workplace* PFs have been exploited by PTs.

Insight 14: The personalization PT has exploited most PFs, but the *social psychology* PFs (especially, AUTHORITY and CONSISTENCY) have been mostly widely exploited by PTs.

B. Which Attacks Exploit Which PTs and PFs?

Which Attacks Exploit Which PTs? We systematize these relationships with respect to the attack types, namely, email-based versus website-based versus OSN-based attacks.

First, email-based attacks have exploited the following PTs.

1) The generic phishing attack has exploited the urgency, attention grabbing, visual deception, incentive and motivator, persuasion, and quid-pro-quo PTs.
2) The spear phishing attack has exploited the urgency, visual deception, incentive and motivator, persuasion, quid-pro-quo, contextualization, pretexting, and personalization PTs.
3) The clone phishing attack has exploited the urgency, attention grabbing, visual deception, incentive and motivator, persuasion, trusted relationship, impersonation, pretexting, and personalization PTs.
4) The whaling attack has exploited the urgency, attention grabbing, visual deception, and personalization PTs.
5) The wire transfer scam attack has exploited the urgency, incentive and motivator, and impersonation PTs.

6) The BEC attack has exploited the urgency, attention grabbing, visual deception, trusted relationship, and impersonation PTs. That is, all these attacks exploit multiple PTs, which explains why they are hard to defend against, and the three phishing attacks have exploited most PTs, which suggests the large attack effort and the large (if not the largest) payback to attackers. In total, these attacks have used 11 PTs, and through which 42 PFs.

Second, website-based attacks have exploited the following PTs: 1) the scareware attack has used the urgency, attention grabbing, quid-pro-quo, and incentive and motivator PTs; 2) the typosquatting attack has used the visual deception PT; 3) the spamdexing attack has used the attention grabbing and trusted relationship PTs; 4) the drive-by download attack has exploited the visual deception and trusted relationship PTs; 5) the click-baiting attack has exploited the visual deception and persuasion PTs; 6) the malvertising attack has exploited the attention grabbing, visual deception, and incentive and motivator PTs; 7) the reverse SE attack has exploited the incentive and motivator, impersonation, and pretexting PTs; 8) the angler phishing attack has exploited the trusted relationship, impersonation, and priming PTs; 9) the pharming attack has exploited the trusted relationship and contextualization attacks; 10) the water holing attack has exploited the trusted relationship PT; 11) the tabnabbing attack has exploited the visual deception and impersonation PTs; and 12) the ad fraud attack has exploited the attention grabbing, visual deception, incentive and motivator, and persuasion PTs. That is, most attacks exploit multiple PTs, which may explain why it is hard to defend against these attacks. In total, these attacks have exploited 13 PTs, and through which 45 PFs.

Third, OSN-based attacks have exploited the following PTs: 1) the honey trap attack has exploited the foot-in-the-door, impersonation, pretexting, personalization, affection trust, and priming PTs; 2) the catfishing attack has exploited the impersonation, affection trust, and priming PTs; 3) the app spoofing attack has exploited the visual deception, impersonation, and loss aversion PTs; 4) the likejacking attack has exploited the visual deception, incentive and motivator, persuasion, and decoy effect PTs; 5) the vishing attack has exploited the urgency, incentive and motivator, persuasion, impersonation, contextualization, pretexting, personalization, and decoy effect PTs; 6) the smishing attack has exploited the urgency, incentive and motivator, persuasion, impersonation, contextualization, pretexting, personalization, decoy effect, and priming PTs; 7) the QRishing attack has exploited the visual deception and trusted relationship PTs; and 8) the Sybil attack has exploited the impersonation and trusted relationship PTs. That is, all these attacks have used multiple PTs, which may explain why they are difficult to defend against. In total, these attacks have exploited nine PTs, and through which 31 PFs.

Which Attacks Exploit Which PFs? Each attack exploits some PF(s) through one or multiple PTs. For example, the spear phishing attack exploits: 1) PFs such as IMPULSIVITY, RESPECT, and SELF-CONTROL through the personalization PT; 2) PFs such as FEAR, RESPECT, SCARCITY, and TRUST through the impersonation PT; and 3) PFs such as COGNITIVE MISER, FEAR, HOPELESSNESS, and NEGLIGENCE through the urgency PT. As highlighted in Fig. 9, one attack can exploit multiple PTs, one PT can be exploited by multiple attacks, one PT can exploit multiple PFs, and one PF can be exploited by multiple PTs; as a result, one attack can exploit multiple PFs, and one PF can be exploited by multiple attacks. We observe that attackers are very aggressive in exploiting PFs. The spearing phishing attack is the most sophisticated in the sense that it exploits 41 PFs through eight PTs.

Insight 15: BEC exploits most PTs and spear phishing exploits most PFs.

C. Which Defenses Leverage Which PTs and PFs?

Which Defenses Exploit Which PTs? Our understanding of leveraging PTs to design effective defenses is even more superficial than our understanding of leveraging PFs to design effective defenses. Nevertheless, we observe that it may be possible to design defenses to leverage PTs, such as by defining features to represent these PTs for machine learning-based defenses, through which PFs can be indirectly leveraged by defenses. For example, the honey trap attack exploits the LONELINESS PF, which may be hard to leverage by machine learning-based defenses. To address this attack, an effective defense should consider a PT that exploits LONELINESS, such as the incentives & motivators PT. Similarly, to address an attack such as the drive-by download, an effective defense should consider the PTs that are leveraged by this attack, such as visual deception and trusted relationship.

Which Defenses Exploit Which PFs? As discussed in Section VIII, there are only four defenses against email-based attacks that have collectively leveraged the following ten PFs in total: COGNITIVE MISER, ABSENT-MINDEDNESS, GREED, FEAR, AUTHORITY, RECIPROCATION, RESPECT, NEGLIGENCE, TRUST, and CURIOSITY by defining features to reflect these PFs for machine learning purposes. There are only five defenses against website-based attacks that have collectively leveraged nine PFs in total: EXPERTISE, OVERCONFIDENCE, ABSENTMINDEDNESS, AUTHORITY, RESPECT, TRUST, VIGILANCE, HABITUATION, and SELF-EFFICACY, also by defining features to reflect these PFs for machine learning purposes. There are only three defenses against OSN-based attacks that have collectively leveraged nine PFs in total: AUTHORITY, RECIPROCATION, LIKING, SCARCITY, SOCIAL PROOF, CONSISTENCY, RESPECT, TRUST, and HABITUATION PFs by defining features to reflect these PFs for machine learning purposes.

Adding the numbers together, we observe that there are 19 PFs in total that have been exploited by defenses,

which is substantially fewer than the 46 PFs that have been exploited by attackers. This may explain why existing defenses have achieved limited success. A further examination shows that some PFs cannot be directly leveraged by defenses because they are inherent to human behaviors, such as IMPATIENCE, IMPULSIVITY, and CURIOSITY. This makes it unclear how to leverage them for machine learning techniques because these factors are not reflected in the attack content (i.e., email, webpage, or OSN content). Nevertheless, this highlights the importance of seeking a more comprehensive framework, which serves as one motivation for the framework that we will present in Section X. For example, it would be important to seek human training-based defenses that adequately incorporate PFs, emphasizing some specific factors that have been deemed important in the literature (e.g., SELF-EFFICACY, EXPERTISE, and HABITUATION [94], [165]). Still, there are open problems, such as understanding and incorporating PFs such as IMPATIENCE or IMPULSIVITY into defenses.

Insight 16: The hybrid phishing detection leverages most PTs, and the Phishpedia defense leverages most PFs, but much research remains to be done on how to leverage PTs and PFs to design effective defenses. This is especially true for PTs and PFs that may be difficult to leverage (e.g., the affection trust and quid-pro-quo PTs and the TRUST, IMPULSIVITY, CURIOSITY, FEAR, and NEGLIGENCE PFs).

D. Which Defenses Address Which Attacks?

Defenses leveraging PFs have been proposed to defend against the following email-based attacks: generic phishing, spear phishing, wire transfer scams, and BEC. Defenses leveraging PFs have been proposed to defend against the following website-based attacks: scareware, typosquatting, malvertising, and angler phishing.

Defenses leveraging PFs have been proposed to defend against the following OSN-based attacks: catfishing, vishing, smishing, and Sybil attacks. We observe that most attacks have not been addressed with defenses that leverage PFs or PTs, further suggesting the ineffectiveness of current defenses.

Insight 17: Many attacks have not been addressed with defenses that leverage PFs or PTs.

X. ROADMAP FOR FUTURE RESEARCH

The preceding exploration prompts us to propose the following roadmap for future research, including an envisioned framework, and applying it to guide the design of innovative defenses. The roadmap has three components: 1) characterizing SE attacks via other psychological lenses (other than the PF and PT lens used in this article); 2) synergizing views from different psychological lenses into a holistic framework; and 3) leveraging the framework to guide the design of effective defenses.

A. Characterizing SE Attacks via Other Psychological Lenses

This study uses the BFPTs and Cialdini's Principles of Persuasion as a starting point to identify PFs and PTs. As shown above, this PF- and PT-centric psychological lens indeed offers a fruitful way to understand humans' vulnerabilities to SE attacks. However, we should stress that there may be other psychological lenses that can be leveraged to conduct studies from complementary perspectives. This offers a good opportunity for future research because a holistic understanding of SE attacks would require us to understand all these perspectives. To see the feasibility, in what follows, we outline one such perspective.

One perspective, which is complementary to the one used in this study, is to leverage the Theory of System 1 (heuristic) versus System 2 (analytic) Information Processing [240] because SE message processing is affected by environmental factors and a variety of cognitive mechanisms. System 1 is based on heuristics and is, thus, fast and effortless, but error-prone; on the other hand, System 2 involves deep analytical thinking and, therefore, is slow and effortful [240]. That is, this theory suggests that deliberate reasoning, typically logical or mathematical, falls under System 2 [241]. Of course, this theory does not come without criticism, despite studies supporting it (e.g., [79], [81], [95], [242], [243], and [244]). This is perhaps because although the characteristics of the two processes are often clear, the factors that determine when an individual will think analytically or rely on their intuition is unclear [245]. It is clear that resolving this problem will have an important application to coping with SE attacks: we can leverage these factors to encourage users to trigger System 2 when dealing with SE attacks. This indeed has inspired some studies in cognitive psychology [246], such as using electroencephalography (EEG) to decipher the underlying neural mechanisms [244] and improve human decision-making capabilities [247]. Moreover, there are attempts to refine this dual-thinking process as a three-stage dual process model [245]. Related to this perspective, a recent study shows that humans can actually process deliberate reasoning involving logical principles in an intuitive fashion (i.e., without deliberation) [241].

B. Synergizing Views From Different Psychological Lenses Into a Comprehensive Framework

We envision that the different views seen through the different psychological lenses can be incorporated and synergized into a unified framework. To see this possibility, in what follows, we discuss how the view seen through our PF-centric lens can be synergized with the view seen through the Theory of System 1 versus System 2 lens into a more comprehensive framework, such as the one outlined in the following.

1) Seeking a Comprehensive Qualitative Framework: Our premise is that in the context of SE attacks, it may

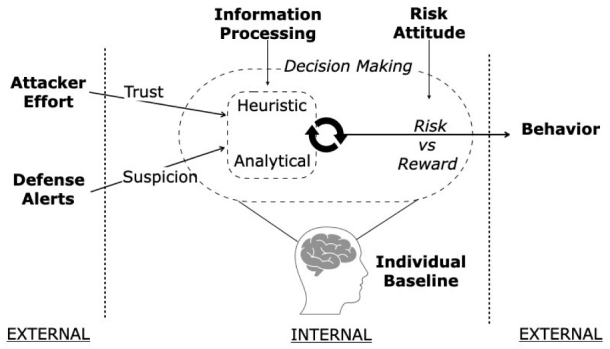


Fig. 10. ***Envisioned qualitative psychological framework for understanding SE attacks.***

not be accurate to regard heuristics as especially error-prone, and thus, both heuristics and analytic processing can help prevent victimization under different conditions. This prompts us to propose a qualitative framework for applying psychology to the cybersecurity domain, which would ultimately lead to a new discipline that may be called *psychological cybersecurity*.

As highlighted in Fig. 10, the framework is based on the Theory of System 1 versus System 2 for describing human information processing of input associated with SE attacks. The framework has five components: *information processing*, *risk attitude*, *individual baseline*, *attacker effort*, and *defense alerts*. These five components collectively determine one's *behavior* (e.g., clicking a link in a phishing email or not). In what follows, we elaborate on these components while discussing how PFs and PTs (obtained through the lens used in this study) will be incorporated.

Information Processing (Heuristic versus Analytic) This component is inspired by the Theory of System 1 versus System 2. It aims to identify the conditions that channel one into *heuristic processing* or *analytic processing*.

1) *Heuristic processing:* This uses patterns and rules, or a trial-and-error approach, to reach a decision. Heuristics are often used in uncertain situations where information or time is limited. In SE attacks, nonexperts are more likely to rely on heuristics to determine the credibility of a message. Other than rules and patterns, nonexpert users also rely on previous experiences (often acquired via trial and error) to detect SE messages [105], [248]. Heuristics are useful but can be erroneous when the rules used to determine credibility are based on elements that can be manipulated by attackers or when they are unable to discriminate between benign and SE messages.
2) *Analytic processing:* This involves evaluating multiple factors to reach a decision. It requires an individual to be knowledgeable about factors relevant to the outcome and have the information required to support a decision. In SE attacks, individuals with cybersecurity EXPERTISE, which is a PF, are more likely to use analytic processing to detect SE attacks (e.g., experts would consider multiple factors before determining the credibility of emails [249]).

The preceding categorization is important because attackers often attempt to deceive victims into heuristic processing over analytic processing to increase their chances of success. This prompts us to propose four core research problems associated with this component: 1) how should heuristic processing and analytic processing work together when dealing with SE attacks? 2) which PFs and/or PTs would force humans to be trapped in System 1 or heuristic processing? 3) how do PFs and/or PTs force humans to be trapped in System 1 or heuristic processing? and 4) is there a threshold of PFs and/or PTs above which a human will use System 2 rather than System 1? Along these lines, there are some initial attempts. For example, it seems that when persuasion is not detected, one would use heuristic reasoning [51], [250], and one would use a negative response otherwise [251].

Risk Attitude: This component is important because studies show that risk attitude affects the likelihood of SE victimization, perhaps even independent of human information processing [88], [153]. This is possible because risk attitude affects motivators, which drive humans to act [252]. There are three risk attitudes: *risk-seeking*, *risk-aversion*, and *risk-neutral*. For example, even when information processing triggers suspicions, a risk-seeking user may still comply with a malicious request because the prospect reward exceeds the perceived risk, explaining: 1) why some people make risky decisions in cyberspace especially when they feel they have little to lose [153], [253] and 2) why some people still fall victim even if they recognize the risk [88]. This prompts us to propose the following research problem: how does risk attribute, perhaps together with PFs and/or PTs, affect a human's decision in applying heuristic versus analytic processing?

Individual Baseline: This component deals with the PFs that may be exploited by attackers. For example, the following PFs may encourage the use of heuristic processing, HABITUATION, STRESS, and WORKLOAD, because HABITUATION reduces suspicions and a combination of STRESS and WORKLOAD may increase the reliance on heuristic processing and decrease VIGILANCE. This prompts us to propose a core research problem associated with this component: how should we characterize the impact of PFs and PTs on humans' susceptibility to SE attacks? This pertains closely to the psychological lens used in this study.

Attacker Effort: The component deals with the external attacker's effort at exploiting PFs to earn victims' trust and encourage their compliance (e.g., how real a phishing email looks). For example, an attacker can earn the trust of a victim by creating emails of high quality and appealing to the victim. This is because many users judge credibility based on superficial attributes, such as

the professional appearance of a website, the absence of grammatical errors, or recognizable logos in emails [166], [254]. To generate an appealing message, an attacker can exploit a combination of PTs (e.g., persuasion, personalization, and contextualization). This prompts us to propose one important research problem: how does the attacker effort, including the PTs and thus PFs it exploits, influence a human's decision to apply the heuristic versus analytic process? Along this direction, a very recent study (by some of the authors) initiates the investigation on how to quantify the psychological sophistication of malicious emails [42] while leveraging some of the PFs and PTs described in this article.

Defense Alerts: This component deals with the alerts provided by the employed defense mechanisms (e.g., machine learning-based detectors or warning systems) to warn users of potential threats and trigger their, for example, VIGILANCE. Intuitively, an effective alert would cause users to switch their attention to the warning information and maintain their attention long enough. An effective warning would trigger suspicion [44], such as cue salience triggering attention switching [255]. This prompts us to propose one important research problem associated with this component: how can defense alerts leverage PFs to offset the influence of attackers?

2) Turning the Envisioned Qualitative Framework Into a Quantitative Framework: The qualitative framework outlined above, or a refined version of it, should be enriched or enhanced to incorporate quantitative metrics and characteristics of SE attacks and defenses. At a broad level, we propose centering quantitative characteristics at the notion of individuals' `susceptibility` to SE attacks as follows:

$$\begin{aligned}&\texttt{susceptibility}\\&= f\,(\text{processing\_route}, \text{risk\_attitude}\\&\qquad \text{individual\_baseline}, \text{attacker\_effort}, \text{defense\_alerts})\end{aligned}$$

where f is a family of mathematical functions that are to be identified by future studies (e.g., via experiments), processing_route means the use of heuristic or analytic processing, and the other four arguments are as described above. This formalism serves as a starting point to answer a range of important research questions quantitatively.

1) How important is the role played by processing_route? Quantitative characterization of this helps answer even deeper questions, such as can we reduce the susceptibility to below a threshold without forcing humans to be trapped into System 2? Being able to answer this question will allow us to design cost-effective, if not optimal defenses because we do not have to force individuals to use System 2 to deal with SE attacks.
2) What is the role exactly played by risk_attitude? A quantitative characterization of this provides immediate guidance in designing defenses because we may have to use different defenses to prevent individuals of different risk_attitude from falling victim to SE attacks.
3) What is the role played by individual_base (e.g., which PFs and PTs contribute most to humans' `susceptibility` to SE attacks)? A quantitative characterization of this will offer guidance in designing cost-effective, if not optimal, defense mechanisms. For example, if TRUST turns out to be an important factor, then researchers should seed defenses to minimize humans' TRUST (e.g., by making people practice zero-trust on everything coming from the Internet). Although there are experimental results showing that PFs and PTs influence humans' susceptibility to SE attacks [62], [256], [257], there are no adequate quantitative characterizations. For example, one study [81] reports that participants identify the HURRY PF and the urgency PT that cause them to fall victim to an attack; another study [34] involving expert interviews reports that participants with high AGREEABLENESS and EXTROVERSION PFs are more susceptible to SE attacks than participants with low scores in these two PFs. This status quo highlights the importance of quantifying the impact of each PF and each PT, as well as the impact of a combination of some PFs and/or PTs.
4) What is the impact of attacker_effort? Intuitively, answering this question would require us to identify the PFs and PTs that are involved in an SE attack. Along this line, the first step is to quantify the psychological sophistication of an SE attack as reflected by the PFs and PTs exhibited by the attack [42].
5) How effective are defense_alerts in reducing humans' `susceptibility` to SE attacks? If they turn out to be truly effective, then researchers should investigate how to make warnings as effective as possible (e.g., using dynamic warnings instead of static warnings in order to reduce HABITUATION [154]).

The framework outlined above will have a broader impact than merely dealing with SE attacks, which are only one kind of cyberattack. This is so because the framework can be incorporated into any holistic framework that aims to model different kinds of cyberattacks. Along this line, we envision that the framework can be seamlessly incorporated into the *cybersecurity dynamics* framework [258], [259], [260] that aims to quantify cybersecurity from a holistic perspective [261], [262], [263]. Indeed, the cybersecurity dynamics framework has already accommodated human susceptibility to SE attacks in preventive and reactive cyber defense dynamics models [264], [265], [266], [267], [268], [269]. However, human susceptibility remains to be incorporated into other kinds of models, such as adaptive, proactive, and active cyber defense dynamics [270], [271], [272], [273], [274]. Since humans exhibiting similar PFs may exhibit a similar degree of

susceptibility to SE attacks, this kind of *dependence* also needs to be adequately considered in holistic cybersecurity models; otherwise, the resulting models may offer misleading results as shown in several families of cybersecurity dynamics models [156], [271], [275], [276], [277], [278].

C. Leveraging the Framework to Guide the Design of Effective Defenses

As hinted above, the deep understanding resulting from the quantitative framework can guide the design of cost-effective, if not optimal, defenses. We propose two complementary classes of defenses: *training*, which aims to keep humans in the loop (i.e., the last line of defense) and is primarily geared toward the processing_route and risk_attitude in the framework, and *automated defenses*, which aims to detect and possibly block SE attacks without relying on human participation and is primarily geared toward the individual_baseline and attack_effort.

1) Designing Effective Training Schemes: Training is of fundamental importance because *humans are the last line of defense* in the sense that they can ultimately decide, for example, whether to click a link in an incoming email or whether to accept the advice that an incoming email is malicious. With respect to processing_route, future research should design effective training schemes to encourage individuals to use System 2, rather than System 1, in their decision-making process when coping with potential SE attacks. This is true until we are certain that humans can indeed process deliberate reasoning involving logical principles in an intuitive fashion (i.e., without deliberation), as reported in [241].

We propose a systematic approach to training, with three levels of increasing advancement: *awareness training*, *intermediate training*, and *advanced training*.

Awareness Training: This is to make the general public aware of the threats of SE attacks by showing them examples of malicious emails. This is important because everyone can be a target of SE attacks and a high awareness may lead to a better posture against SE attacks [15], [16]. We propose that an effective awareness training scheme should strive to make people aware the following: 1) cyberspace is a dangerous place where attackers exploit every opportunity to attack innocent people; 2) a reply to an email one never wrote is likely a malicious email; 3) one should not provide any secret information, such as their social security number, in any session unless they initiated the session that requires it; 4) an email or message stating that one won a lottery or game that they did not play is likely malicious; 5) any too-good-to-be-true deal that one did not ask for is likely a bait to entice recipients; 6) being asked for a wire transfer to pay a bill (e.g., utility and cable) that one usually pays with credit card or direct payment is likely a scam; 7) do not assume that automated defenses (e.g., spam filters) are perfect, as they can never be; 8) when something does not feel right, pay the due diligence and double check; 9) being asked to send a gift card number to pay for something, especially a recurrent bill, is likely a scam; 10) any email whose subject line is an email address is likely impersonating a legitimate entity; and 11) click on the top links returned by a search engine with caution because they may be maliciously placed by attackers exploiting search engine optimization.

Intermediate Training: This aims to train people with more in-depth knowledge and skills in dealing with SE attacks. This is suitable for organizations to train employees because any successful SE attack against an employee can cause great damage; for example, BEC attacks usually target financial departments and cause US$132 559 damages on average [3]. Thus, organizations should make this training, which has been practiced by many, if not all, organizations, mandatory. This training should be periodically updated by leveraging the new trends in SE attacks. One aspect that could make training more effective is to demonstrate the potential damages that can be incurred to the organization if they fall victim to an SE attack.

Advanced Training: This is to train employees with advanced skills. It is important, especially for employees with critical job duties, such as the employees in the finance industry and the information technology department. This training should be constantly updated by leveraging the new trends in SE attacks. This training may be conducted in a game environment where employees (trainees) are presented with hypothetical or emulated SE situations and are asked to decide what they should do. This training can further leverage the new SE attacks that have been identified or proposed by researchers but have not been seen in the wild.

2) Designing Automated Defenses: Automated defenses are indispensable because of the amount of SE attacks. For example, APWG reports that there are 1 350 037 phishing attacks in the fourth quarter of 2022 [3]. Therefore, we must leverage automated solutions to reduce the number of malicious emails that need to be analyzed by humans to determine whether they are legitimate or not because of the false negatives of automated solutions. Specifically, future research should adequately leverage the quantitative characterization results obtained in the framework outlined above. With respect to individual_baseline, future defenses should adequately leverage the PFs and PTs to design machine learning-based defenses to provide (for example) defense alerts. For example, future research must seek feature representations to adequately accommodate PFs and PTs to cope with SE attacks. One challenge is, as suggested by Insight 16, that some PTs and PFs may be difficult to leverage when designing defenses.

With respect to attacker_effort, future research should strive to automatically recognize PFs and PTs used in SE attacks, as heavy use of PFs and PTs would immediately flag an SE attack. To the best of our knowledge, automatically identifying PFs and PTs is an unexplored problem, as a recent study [42] on quantifying the psychological

sophistication of SE attacks relies on manual recognition of PFs and PTs, also shown in this study. Being able to automatically recognize PFs and PTs would pave the way for design defenses.

XI. CONCLUSION

We have presented a systematization of SE attacks through the psychological lens centered on PFs, which are the root causes of the problem, and PTs, which are the strategies that have been used by SE attacks to exploit PFs. We have systematized the SE attacks that have been reported in academic and nonacademic venues, as well as the defenses that have been reported in academic venues. To clearly describe the state of the art, we have presented a mapping between PFs, PTs, SE attacks, and defenses. In addition, we have presented a systematic roadmap toward adequately mitigating SE attacks. The roadmap offers a range of future research directions, especially the envisioned framework that aims at a comprehensive understanding of SE attacks through psychological lenses, which would be more extensive than the specific lens used in this study.

There are many open problems for future research. First, what are the other psychological lenses than the one used in this study (i.e., the PFs and PTs built on top of the BFPTs and Cialdini's Principles of Persuasion) and the one discussed in Section X-A and built on top of the Theory of System 1 versus System 2? Studies based on these and possibly other psychological lenses will help build a more comprehensive understanding of SE attacks from the psychological perspective and pave the way toward creating the afore-envisioned discipline of *psychological cybersecurity*.

Second, with respect to the qualitative framework outlined in Section X-B1, we highlight the most exciting open problems with respect to the framework's five components.

1) With respect to *information processing*, is it possible to effectively mitigate SE attacks *without* forcing individuals to use System 2?
2) With respect to *risk attitude*, what are effective strategies for mitigating the damages associated with risk attitude?
3) With respect to *individual baseline*, which PFs and PTs have a high impact on humans' susceptibility to SE attacks?
4) With respect to *attacker effort*, how can we automatically recognize the exploitation of PFs and PTs in SE attacks?
5) With respect to *defense alerts*, how can we optimize the alerts that must be presented by automated defenses to users who will make a decision on whether there is an SE attack?

There are two challenges: 1) automated defenses must minimize the number of emails that must be presented to users for manual processing and 2) machine learning-based automated defenses can be circumvented by *adversarial examples*, such as adversarial emails, which are crafted to evade them.

Third, with respect to the quantitative framework outlined in Section X-B2, we raised a range of important unsolved problems. In addition, we highlight one more open problem: What would be the optimal combination of PFs (or PTs) that could be exploited by SE attackers to maximize their gain? This corresponds to the worst case scenario from a defender's point of view and would demand a defense that can simultaneously recognize and cope with such combinations.

Fourth, with respect to leveraging the framework to design effective defenses as described in Section X-C, unsolved problems include the design of training schemes and automated defenses. In terms of designing training schemes, intermediate training needs to keep up with the trends of real-world SE attacks, if not leveraging forecast SE attacks that may not have been seen in the wild; advanced training for employees of critical job duties would need to be based on innovative designs, possibly games that can demonstrate what catastrophic damages may be incurred by successful SE attacks. In terms of designing automated defenses, automated defenses should leverage the PFs and PTs that can be exploited by attackers *without* keeping humans in the loop. ■

Acknowledgment

The authors thank Eric Ficke and Shawn Emery for proofreading this article. They also thank the Anti-Phishing Working Group (APWG) for providing malicious emails for their study. They also thank the anonymous reviewers for their comments that guided the authors in revising and improving this article.

REFERENCES

[1] FBI. (Apr. 2020). *Business Email Compromise*. [Online]. Available: https://www.fbi.gov/scams-and-safety/common-scams-and-crimes/business-email-compromise
[2] *Business Email Compromise*, FBI, Washington, DC, USA, May 2022.
[3] *Phishing Activity Trends Report—Unifying the Global Response to Cybercrime*, Anti-Phishing Working Group (APWG), Lexington, MA, USA, Feb. 2021.
[4] C. Catal, G. Giray, B. Tekinerdogan, S. Kumar, and S. Shukla, "Applications of deep learning for phishing detection: A systematic literature review," *Knowl. Inf. Syst.*, vol. 64, no. 6, pp. 1457–1500, Jun. 2022.
[5] F. Salahdine and N. Kaabouch, "Social engineering attacks: A survey," *Future Internet*, vol. 11, no. 4, p. 89, Apr. 2019.
[6] S. Asiri, Y. Xiao, S. Alzahrani, S. Li, and T. Li, "A survey of intelligent detection designs of HTML URL phishing attacks," *IEEE Access*, vol. 11, pp. 6421–6443, 2023.
[7] R. Zieni, L. Massari, and M. C. Calzarossa, "Phishing or not phishing? A survey on the detection of phishing websites," *IEEE Access*, vol. 11, pp. 18499–18519, 2023.
[8] G. Petrič and K. Roer, "The impact of formal and informal organizational norms on susceptibility to phishing: Combining survey and field experiment data," *Telematics Informat.*, vol. 67, Feb. 2022, Art. no. 101766.
[9] W. Syafitri, Z. Shukur, U. A. Mokhtar, R. Sulaiman, and M. A. Ibrahim, "Social engineering attacks prevention: A systematic literature review," *IEEE Access*, vol. 10, pp. 39325–39343, 2022.
[10] W. Fuertes et al., "Impact of social engineering attacks: A literature review," in *Proc. Develop. Adv. Defense Secur. (MICRADS)*, 2021, pp. 25–35.
[11] M. A. Siddiqi, W. Pak, and M. A. Siddiqi, "A study on the psychology of social engineering-based cyberattacks and existing countermeasures," *Appl. Sci.*, vol. 12, no. 12, p. 6042, Jun. 2022.
[12] A. Alharbi, H. Dong, X. Yi, Z. Tari, and I. Khalil, "Social media identity deception detection: A survey," *ACM Comput. Surv.*, vol. 54, no. 3, pp. 1–35, 2021.
[13] A. K. Jain and B. B. Gupta, "A survey of phishing attack techniques, defence mechanisms and open

research challenges," *Enterprise Inf. Syst.*, vol. 16, no. 4, pp. 527–565, Apr. 2022.
[14] L. Tang and Q. H. Mahmoud, "A survey of machine learning-based solutions for phishing website detection," *Mach. Learn. Knowl. Extraction*, vol. 3, no. 3, pp. 672–694, Aug. 2021.
[15] M. Hijji and G. Alam, "A multivocal literature review on growing social engineering based cyber-attacks/threats during the COVID-19 pandemic: Challenges and prospective solutions," *IEEE Access*, vol. 9, pp. 7152–7169, 2021.
[16] N. Mashtalyar, U. N. Ntaganzwa, T. Santos, S. Hakak, and S. Ray, "Social engineering attacks: Recent advances and challenges," in *Proc. Int. Conf. Human-Comput. Interact.* Cham, Switzerland: Springer, 2021 pp. 417–431.
[17] R. Zaimi, M. Hafidi, and M. Lamia, "Survey paper: Taxonomy of website anti-phishing solutions," in *Proc. 7th Int. Conf. Social Netw. Anal., Manage. Secur. (SNAMS)*. Paris, France: IEEE, Dec. 2020, pp. 1–8.
[18] M. M. Ali, M. S. Qaseem, and M. A. U. Rahman, "A survey on deceptive phishing attacks in social networking environments," in *Proc. 3rd Int. Conf. Comput. Intell. Inform.* Delhi, India: Springer, 2020, pp. 443–452.
[19] C. M. R. D. Silva, E. L. Feitosa, and V. C. Garcia, "Heuristic-based strategy for phishing prediction: A survey of URL-based approach," *Comput. Secur.*, vol. 88, Jan. 2020, Art. no. 101613.
[20] R. Alabdan, "Phishing attacks survey: Types, vectors, and technical approaches," *Future Internet*, vol. 12, no. 10, p. 168, Sep. 2020.
[21] M. Vijayalakshmi, S. Mercy Shalinie, M. H. Yang, and U. R. U. Meenakshi, "Web phishing detection techniques: A survey on the state-of-the-art, taxonomy and future directions," *IET Netw.*, vol. 9, no. 5, pp. 235–246, Sep. 2020.
[22] D. Jampen, G. Gür, T. Sutter, and B. Tellenbach, "Don't click: Towards an effective anti-phishing training. A comparative literature review," *Human-Centric Comput. Inf. Sci.*, vol. 10, no. 1, pp. 1–41, Dec. 2020.
[23] S. Chanti and T. Chithralekha, "Classification of anti-phishing solutions," *Social Netw. Comput. Sci.*, vol. 1, no. 1, pp. 1–18, Jan. 2020.
[24] J. Rastenis, S. Ramanauskaite, J. Janulevičius, A. Čenys, A. Slotkiene, and K. Pakrijauskas, "E-mail-based phishing attack taxonomy," *Appl. Sci.*, vol. 10, no. 7, p. 2363, Mar. 2020.
[25] D. N. Alharthi, M. M. Hammad, and A. C. Regan, "A taxonomy of social engineering defense mechanisms," in *Proc. Future Inf. Commun. Conf.* Vancouver, BC, Canada: Springer, 2020, pp. 27–41.
[26] A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, and K. Kifayat, "A comprehensive survey of AI-enabled phishing attacks detection techniques," *Telecommun. Syst.*, vol. 76, no. 1, pp. 139–154, Jan. 2021.
[27] Z. Guo, J.-H. Cho, I.-R. Chen, S. Sengupta, M. Hong, and T. Mitra, "Online social deception and its countermeasures: A survey," *IEEE Access*, vol. 9, pp. 1770–1806, 2021.
[28] A. Das, S. Baki, A. El Aassal, R. Verma, and A. Dunbar, "SoK: A comprehensive reexamination of phishing research from the security perspective," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 671–708, 1st Quart., 2020.
[29] A. Yasin, R. Fatima, L. Liu, A. Yasin, and J. Wang, "Contemplating social engineering studies and attack scenarios: A review study," *Secur. Privacy*, vol. 2, no. 4, p. e73, Jul. 2019.
[30] S. R. Sahoo and B. B. Gupta, "Classification of various attacks and their defence mechanism in online social networks: A survey," *Enterprise Inf. Syst.*, vol. 13, no. 6, pp. 832–864, Jul. 2019.
[31] B. B. Gupta, N. A. G. Arachchilage, and K. E. Psannis, "Defending against phishing attacks: Taxonomy of methods, current issues and future directions," *Telecommun. Syst.*, vol. 67, no. 2, pp. 247–267, Feb. 2018.
[32] K. L. Chiew, K. S. C. Yong, and C. L. Tan, "A survey of phishing attacks: Their types, vectors and technical approaches," *Expert Syst. Appl.*, vol. 106, pp. 1–20, Sep. 2018.
[33] K. Huang, M. Siegel, and S. Madnick, "Systematically understanding the cyber attack business: A survey," *ACM Comput. Surveys*, vol. 51, no. 4, pp. 1–36, Jul. 2019.
[34] B. Cusack and K. Adedokun, "The impact of personality traits on user's susceptibility to social engineering attacks," in *Proc. 16th Austral. Inf. Secur. Manag. Conf.*, Perth, WA, Australia, 2018, pp. 83–89.
[35] Z. Dou, I. Khalil, A. Khreishah, A. Al-Fuqaha, and M. Guizani, "Systematization of knowledge (SoK): A systematic review of software-based web phishing detection," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2797–2819, 4th Quart., 2017.
[36] A. Aleroud and L. Zhou, "Phishing environments, techniques, and countermeasures: A survey," *Comput. Secur.*, vol. 68, pp. 160–196, Jul. 2017.
[37] S. Gupta, A. Singhal, and A. Kapoor, "A literature survey on social engineering attacks: Phishing attack," in *Proc. Int. Conf. Comput., Commun. Autom. (ICCCA)*. Greater Noida, India: IEEE, Apr. 2016, pp. 537–540.
[38] R. Heartfield and G. Loukas, "A taxonomy of attacks and a survey of defence mechanisms for semantic social engineering attacks," *ACM Comput. Surv.*, vol. 48, no. 3, pp. 1–39, Feb. 2016.
[39] K. RaniSahu and J. Dubey, "A survey on phishing attacks," *Int. J. Comput. Appl.*, vol. 88, no. 10, pp. 42–45, Feb. 2014.
[40] M. Khonji, Y. Iraqi, and A. Jones, "Phishing detection: A literature survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 2091–2121, 4th Quart, 2013.
[41] A. Almomani, B. B. Gupta, S. Atawneh, A. Meulenberg, and E. Almomani, "A survey of phishing email filtering techniques," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 2070–2090, 4th Quart., 2013.
[42] R. Montañez et al., "Quantifying psychological sophistication of malicious emails," in *Proc. Int. Conf. Sci. Cyber Secur.* Cham, Switzerland: Springer, 2023, pp. 319–331.
[43] P. Burda, L. Allodi, and N. Zannone, "Dissecting social engineering attacks through the lenses of cognition," in *Proc. IEEE Eur. Symp. Secur. Privacy Workshops (EuroS&PW)*, Sep. 2021, pp. 149–160.
[44] R. Montañez, E. Golob, and S. Xu, "Human cognition through the lens of social engineering cyberattacks," *Frontiers Psychol.*, vol. 11, pp. 1–18, Sep. 2020, Art. no. 1755.
[45] S. Uebelacker and S. Quiel, "The social engineering personality framework," in *Proc. Workshop Socio-Tech. Aspects Secur. Trust*, Jul. 2014, pp. 24–30.
[46] P. Tetri and J. Vuorinen, "Dissecting social engineering," *Behaviour Inf. Technol.*, vol. 32, no. 10, pp. 1014–1023, Oct. 2013.
[47] P. Dolan, M. Hallsworth, D. Halpern, D. King, and I. Vlaev, "Mindspace: Influencing behaviour for public policy," Inst. Government, London, U.K., Tech. Rep. 35792, 2010.
[48] L. Cranor, "A framework for reasoning about the human in the loop," in *Proc. 1st Conf. Usability, Psychol., Secur. (UPSEC)*. USA: USENIX Association, 2008, pp. 1–15.
[49] R. Montaez, A. Atyabi, and S. Xu, "Social engineering attacks and defenses in the physical world vs. cyberspace: A contrast study," in *Cybersecurity and Cognitive Science*. Philadelphia, PA, USA: Elsevier, 2022, pp. 3–41.
[50] P. Burda, L. Allodi, and N. Zannone, "Cognition in social engineering empirical research: A systematic literature review," *ACM Trans. Comput.-Human Interact.*, vol. 31, no. 2, pp. 1–55, Apr. 2024.
[51] R. B. Cialdini and L. James, *Influence: Science and Practice*, vol. 4. Boston, Boston, MA, USA: Pearson Education, 2009.
[52] T. Lin et al., "Susceptibility to spear-phishing emails: Effects of Internet user demographics and email content," *ACM Trans. Comput.-Human Interact.*, vol. 26, no. 5, pp. 1–28, Oct. 2019.
[53] P. Rajivan and C. Gonzalez, "Creative persuasion: A study on adversarial behaviors and strategies in phishing attacks," *Frontiers Psychol.*, vol. 9, p. 135, Feb. 2018.
[54] A. Ferreira and G. Lenzini, "An analysis of social engineering principles in effective phishing," in *Proc. Workshop Socio-Technical Aspects Secur. Trust.*, Jul. 2015, pp. 9–16.
[55] F. Stajano and P. Wilson, "Understanding scam victims: Seven principles for systems security," *Commun. ACM*, vol. 54, no. 3, pp. 70–75, Mar. 2011.
[56] A. van der Heijden and L. Allodi, "Cognitive triaging of phishing attacks," in *Proc. 28th USENIX Security Symp.* Santa Clara, CA, USA: USENIX, 2019, pp. 1309–1326.
[57] A. Ferreira, L. Coventry, and G. Lenzini, "Principles of persuasion in social engineering and their use in phishing," in *Proc. 3rd Int. Conf. Human Aspects Inf. Secur., Privacy, Trust (HAS)*, Los Angeles, CA, USA. Cham, Switzerland: Springer, Aug. 2015, pp. 36–47.
[58] R. B. Cialdini and N. J. Goldstein, "The science and practice of persuasion," *Cornell Hotel Restaurant Admin. Quart.*, vol. 43, no. 2, pp. 40–50, 2002.
[59] J. M. Hatfield, "Social engineering in cybersecurity: The evolution of a concept," *Comput. Secur.*, vol. 73, pp. 102–113, Mar. 2018.
[60] J. H. Bullée, L. Montoya, W. Pieters, M. Junger, and P. Hartel, "On the anatomy of social engineering attacks—A literature-based dissection of successful attacks," *J. Investigative Psychol. Offender Profiling*, vol. 15, no. 1, pp. 20–45, Jan. 2018.
[61] S. Das, A. D. Kramer, L. A. Dabbish, and J. I. Hong, "Increasing security sensitivity with social proof: A large-scale experimental confirmation," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.* New York, NY, USA: ACM, 2014, pp. 739–749.
[62] A. Algarni, Y. Xu, and T. Chan, "An empirical study on the susceptibility to social engineering in social networking sites: The case of Facebook," *Eur. J. Inf. Syst.*, vol. 26, no. 6, pp. 661–687, Nov. 2017.
[63] Y. Kano and T. Nakajima, "Trust factors of social engineering attacks on social networking services," in *Proc. IEEE 3rd Global Conf. Life Sci. Technol. (LifeTech)*. Osaka, Japan: IEEE, Mar. 2021, pp. 25–28.
[64] A. Ferreira, "Why ransomware needs a human touch," in *Proc. Int. Carnahan Conf. Secur. Technol. (ICCST)*. Quebec, QC, Canada: IEEE, Oct. 2018, pp. 1–5.
[65] E. D. Frauenstein and S. Flowerday, "Susceptibility to phishing on social network sites: A personality information processing model," *Comput. Secur.*, vol. 94, Jul. 2020, Art. no. 101862.
[66] L. R. Goldberg, "Language and individual differences: The search for universals in personality lexicons," *Rev. Personality Social Psychol.*, vol. 2, no. 1, pp. 141–165, 1981.
[67] B. W. Roberts, N. R. Kuncel, R. Shiner, A. Caspi, and L. R. Goldberg, "The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes," *Perspect. Psychol. Sci.*, vol. 2, no. 4, pp. 313–345, Dec. 2007.
[68] P. T. Costa Jr. and R. R. McCrae, "The revised neo personality inventory (NEO-PI-R)," in *The SAGE Handbook of Personality Theory and Assessment*. Newbury Park, CA, USA: Sage, 2008.
[69] J. M. Digman, "Personality structure: Emergence of the five-factor model," *Annu. Rev. Psychol.*, vol. 41, no. 1, pp. 417–440, Jan. 1990.
[70] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications," *J. Personality*, vol. 60, no. 2, pp. 175–215, Jun. 1992.
[71] C. J. Soto, "How replicable are links between personality traits and consequential life outcomes? The life outcomes of personality replication project," *Psychol. Sci.*, vol. 30, no. 5, pp. 711–727, May 2019.
[72] D. Nettle, "The evolution of personality variation in humans and other animals," *Amer. Psychologist*,

vol. 61, no. 6, pp. 622–631, 2006.
[73] M. Al-Hamar, R. Dawson, and L. Guan, "A culture of trust threatens security and privacy in Qatar," in *Proc. 10th IEEE Int. Conf. Comput. Inf. Technol.*, Jun. 2010, pp. 991–995.
[74] S. Goel, K. Williams, and E. Dincelli, "Got phished? Internet security and human vulnerability," *J. Assoc. Inf. Syst.*, vol. 18, no. 1, pp. 22–44, Jan. 2017.
[75] V. Garousi, M. Felderer, and M. V. Mäntylä, "Guidelines for including grey literature and conducting multivocal literature reviews in software engineering," *Inf. Softw. Technol.*, vol. 106, pp. 101–121, Feb. 2019.
[76] M. M. Ahmed, "Social engineering attacks in E-government system: Detection and prevention," *Int. J. Appl. Eng. Manage. Lett.*, vol. 6, no. 1, pp. 100–116, Feb. 2022.
[77] A. Fraudwatch. (Apr. 2017). *Angler Phishing: The Risks and Dangers of Fake Social Media Brand Profiles—Part 1*. [Online]. Available: https://fraudwatch.com/angler-phishing-the-risks-and-dangers-of-fake-social-media-brand-profiles-part-1/
[78] J. Reb, J. Narayanan, and Z. W. Ho, "Mindfulness at work: Antecedents and consequences of employee awareness and absent-mindedness," *Mindfulness*, vol. 6, no. 1, pp. 111–122, Feb. 2015.
[79] H. Zafar, A. Randolph, S. Gupta, and C. Hollingsworth, "Traditional SETA no more: Investigating the intersection between cybersecurity and cognitive neuroscience," in *Proc. Annu. Hawaii Int. Conf. Syst. Sci.* Honolulu, HI, USA: University of Hawai'i at Mānoa Hamilton Library, 2019, pp. 4914–4923.
[80] H. Collier and A. Collier, "The port Z3R0 effect! Human behaviors related to susceptibility," *Nature*, vol. 2, no. 3, p. 5, 2020.
[81] N. H. Chowdhury, M. T. P. Adam, and G. Skinner, "The impact of time pressure on cybersecurity behaviour: A systematic literature review," *Behaviour Inf. Technol.*, vol. 38, no. 12, pp. 1290–1308, Dec. 2019.
[82] B. K. Attell, K. Kummerow Brown, and L. A. Treiber, "Workplace bullying, perceived job stressors, and psychological distress: Gender and race differences in the stress process," *Social Sci. Res.*, vol. 65, pp. 210–221, Jul. 2017.
[83] F. Asey, "The wretched of the work: Anger, fear, and hopelessness as impacts of experiencing workplace racism in British Columbia, Canada," *Crit. Social Work*, vol. 22, no. 2, pp. 2–23, Jan. 2022.
[84] P. Tulkarm, "A survey of social engineering attacks: Detection and prevention tools," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 18, pp. 4375–4386, 2021.
[85] Tessian. (Nov. 2023). *15 Examples of Real Social Engineering Attacks—Updated 2023*. [Online]. Available: https://www.tessian.com/blog/examples-of-social-engineering-attacks/
[86] A. Petrosyan. (Nov. 2023). *Industries Most Targeted by Web Application Attacks 2022*. [Online]. Available: https://www.statista.com/statistics/221293/cyber-crime-target-industries/
[87] S. M. Kerner. (Apr. 2022). *Colonial Pipeline Hack Explained: Everything You Need to Know*. [Online]. Available: https://www.techtarget.com/whatis/feature/Colonial-Pipeline-hack-explained-Everything-you-need-to-know
[88] S. E. Lea, P. Fischer, and K. M. Evans, "The psychology of scams: Provoking and committing errors of judgement," Office Fair Trading, London, U.K., Tech. Rep. OFT1070, 2009.
[89] D. Diaz. (Feb. 2023). *What are Personality and Individual Differences?* [Online]. Available: https://online.sunderland.ac.uk/what-are-personality-and-individual-differences
[90] K. Cherry, "The big five personality dimensions: 5 major factors of personality," Tech Republic, Louisville, KY, USA, 2012.
[91] Z. Wang, L. Sun, and H. Zhu, "Defining social engineering in cybersecurity," *IEEE Access*, vol. 8, pp. 85094–85115, 2020.
[92] N. M. Trent, J. W. Joubert, and W. L. Bean, "Engineering students' perspectives on the use of group work peer assessment in two undergraduate industrial engineering modules," in *Proc. World Eng. Educ. Forum-Global Eng. Deans Council (WEEF-GEDC)*, Nov. 2020, pp. 1–5.
[93] C. Nobles, "Establishing human factors programs to mitigate blind spots in cybersecurity," in *Proc. MWAIS*, vol. 22, 2019, pp. 1–6.
[94] H. Tu, A. Doupé, Z. Zhao, and G.-J. Ahn, "Users really do answer telephone scams," in *Proc. 28th USENIX Secur. Symp. (USENIX Security)*. Berkeley, CA, USA: USENIX, 2019, pp. 1327–1340.
[95] P. Schaab, K. Beckers, and S. Pape, "Social engineering defence mechanisms and counteracting training strategies," *Inf. Comput. Secur.*, vol. 25, no. 2, pp. 206–222, Jun. 2017.
[96] E. J. Williams, A. Beardmore, and A. N. Joinson, "Individual differences in susceptibility to online influence: A theoretical review," *Comput. Hum. Behav.*, vol. 72, pp. 412–421, Jul. 2017.
[97] I. Ghafir, V. Prenosil, A. Alhejailan, and M. Hammoudeh, "Social engineering attack strategies and defence approaches," in *Proc. IEEE 4th Int. Conf. Future Internet Things Cloud (FiCloud)*. Vienna, Austria: IEEE, Aug. 2016, pp. 145–149.
[98] G. D. Moody, D. F. Galletta, and B. K. Dunn, "Which phish get caught? An exploratory study of individuals susceptibility to phishing," *Eur. J. Inf. Syst.*, vol. 26, no. 6, pp. 564–584, Nov. 2017.
[99] Z. Wang, H. Zhu, and L. Sun, "Social engineering in cybersecurity: Effect mechanisms, human vulnerabilities and attack methods," *IEEE Access*, vol. 9, pp. 11895–11910, 2021.
[100] H. Aldawood and G. Skinner, "A taxonomy for social engineering attacks via personal devices," *Int. J. Comput. Appl.*, vol. 178, no. 50, pp. 19–26, Sep. 2019.
[101] K. Zheng, T. Wu, X. Wang, B. Wu, and C. Wu, "A session and dialogue-based social engineering framework," *IEEE Access*, vol. 7, pp. 67781–67794, 2019.
[102] M. Workman, "Gaining access with social engineering: An empirical study of the threat," *Inf. Syst. Secur.*, vol. 16, no. 6, pp. 315–331, Dec. 2007.
[103] W. D. Kearney and H. A. Kruger, "Can perceptual differences account for enigmatic information security behaviour in an organisation?" *Comput. Secur.*, vol. 61, pp. 46–58, Aug. 2016.
[104] I. Kirlappos, S. Parkin, and M. A. Sasse, "Learning from 'shadow security': Why understanding non-compliant behaviors provides the basis for effective security," in *Proc. Workshop Usable Secur.*, 2014, pp. 1–10.
[105] E. M. Redmiles, N. Chachra, and B. Waismeyer, "Examining the demand for spam: Who clicks?" in *Proc. CHI Conf. Human Factors Comput. Syst.* Montreal, QC, Canada: ACM, Apr. 2018, p. 212.
[106] N. Abe and M. Soltys, "Deploying health campaign strategies to defend against social engineering threats," *Proc. Comput. Sci.*, vol. 159, pp. 824–831, Jan. 2019.
[107] A. Beghdadi, M. A. Qureshi, S. A. Amirshahi, A. Chetouani, and M. Pedersen, "A critical analysis on perceptual contrast and its use in visual information analysis and processing," *IEEE Access*, vol. 8, pp. 156929–156953, 2020.
[108] K. Schulz and G. U. Hayn-Leichsenring, "Face attractiveness versus artistic beauty in art portraits: A behavioral study," *Frontiers Psychol.*, vol. 8, p. 2254, Dec. 2017.
[109] T. Halevi, N. Memon, and O. Nov, "Spear-phishing in the wild: A real-world study of personality, phishing self-efficacy and vulnerability to spear-phishing attacks," in *SSRN Electron. J.*, Jan. 2015. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2544742
[110] I. Alseadoon, M. Othman, and T. Chan, "What is the influence of users' characteristics on their ability to detect phishing emails?" in *Advanced Computer and Communication Engineering Technology*. Cham, Switzerland: Springer, 2015, pp. 949–962.
[111] D. Yuan et al., "Detecting fake accounts in online social networks at the time of registrations," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.* New York, NY, USA: ACM, Nov. 2019, pp. 1423–1438.
[112] B. Gancherov. (Sep. 2017). *3 Misaligned Incentives That Are Threatening Your Cybersecurity*. [Online]. Available: https://www.dyntek.com/blog/3-misaligned-incentives-that-are-threatening-your-cybersecurity
[113] N. S. Safa, M. Sookhak, R. Von Solms, S. Furnell, N. A. Ghani, and T. Herawan, "Information security conscious care behaviour formation in organizations," *Comput. Secur.*, vol. 53, pp. 65–78, Sep. 2015.
[114] M. Adil, R. Khan, and M. A. Nawaz Ul Ghani, "Preventive techniques of phishing attacks in networks," in *Proc. 3rd Int. Conf. Advancements Comput. Sci. (ICACS)*. Lahore, Pakistan: IEEE, Feb. 2020, pp. 1–8.
[115] T. Li, K. Wang, and J. Horkoff, "Towards effective assessment for social engineering attacks," in *Proc. IEEE 27th Int. Requirements Eng. Conf. (RE)*. Jeju Island, South Korea: IEEE, Sep. 2019, pp. 392–397.
[116] J. D. Ndibwile, E. T. Luhanga, D. Fall, D. Miyamoto, G. Blanc, and Y. Kadobayashi, "An empirical approach to phishing countermeasures through smart glasses and validation agents," *IEEE Access*, vol. 7, pp. 130758–130771, 2019.
[117] J. G. Zheng et al., "Entity linking for biomedical literature," *BMC Med. Informat. Decis. Making*, vol. 15, no. S1, p. S4, Dec. 2015.
[118] S. G. A. van de Weijer and E. R. Leukfeldt, "Big five personality traits of cybercrime victims," *Cyberpsychol., Behav., Social Netw.*, vol. 20, no. 7, pp. 407–412, Jul. 2017.
[119] T. J. Holt, J. van Wilsem, S. van de Weijer, and R. Leukfeldt, "Testing an integrated self-control and routine activities framework to examine malware infection victimization," *Social Sci. Comput. Rev.*, vol. 38, no. 2, pp. 187–206, Apr. 2020.
[120] J.-W. Bullee, L. Montoya, M. Junger, and P. Hartel, "Spear phishing in organisations explained," *Inf. Comput. Secur.*, vol. 25, no. 5, pp. 593–613, Nov. 2017.
[121] A. K. Welk, K. W. Hong, O. A. Zielinska, R. Tembe, E. Murphy-Hill, and C. B. Mayhorn, "Will the 'phisher-men' reel you in? Assessing individual differences in a phishing detection task," *Int. J. Cyber Behav., Psychol. Learn.*, vol. 5, no. 4, pp. 1–17, 2015.
[122] M. T. Whitty, "Do you love me? Psychological characteristics of romance scam victims," *Cyberpsychol., Behav., Social Netw.*, vol. 21, no. 2, pp. 105–109, Feb. 2018.
[123] I. Alseadoon, T. Chan, E. Foo, and J. G. Nieto, "Who is more susceptible to phishing emails? A Saudi Arabian study," in *Proc. 23rd Australas. Conf. Inf. Syst. (ACIS)*. Victoria, SA, Australia: Deakin Univ., 2012, pp. 1–11.
[124] H. Siadati, T. Nguyen, P. Gupta, M. Jakobsson, and N. Memon, "Mind your SMSes: Mitigating social engineering in second factor authentication," *Comput. Secur.*, vol. 65, pp. 14–28, Mar. 2017.
[125] L. Xiangyu, L. Qiuyang, and S. Chandel, "Social engineering and insider threats," in *Proc. Int. Conf. Cyber-Enabled Distrib. Comput. Knowl. Discovery (CyberC)*. Nanjing, China: IEEE, Oct. 2017, pp. 25–34.
[126] P. Wang, X. Liao, Y. Qin, and X. Wang, "Into the deep web: Understanding e-commerce fraud from autonomous chat with cybercriminals," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, 2020, pp. 1–16.
[127] M. S. Jalali, M. Bruckes, D. Westmattelmann, and G. Schewe, "Why employees (still) click on phishing links: Investigation in hospitals," *J. Med. Internet Res.*, vol. 22, no. 1, 2020, Art. no. e16775.
[128] J. McAlaney and V. Benson, "Cybersecurity as a social phenomenon," in *Cyber Influence and Cognitive Threats*. Philadelphia, PA, USA: Elsevier, 2020, pp. 1–8.
[129] Y. Kim and H. Lee, "Towards a sustainable news business: Understanding readers' perceptions of

algorithm-generated news based on cultural conditioning," *Sustainability*, vol. 13, no. 7, p. 3728, Mar. 2021.
[130] J. McAlaney and P. J. Hills, "Understanding phishing email processing and perceived trustworthiness through eye tracking," *Frontiers Psychol.*, vol. 11, p. 1756, Jul. 2020.
[131] S. M. Albladi and G. R. S. Weir, "Predicting individuals' vulnerability to social engineering in social networks," *Cybersecurity*, vol. 3, no. 1, pp. 1–19, Dec. 2020.
[132] S. Das, A. Kim, Z. Tingle, and C. Nippert-Eng, "All about phishing: Exploring user research through a systematic literature review," 2019, *arXiv:1908.05897*.
[133] D. Henshel, M. G. Cains, B. Hoffman, and T. Kelley, "Trust as a human factor in holistic cyber security risk assessment," *Proc. Manuf.*, vol. 3, pp. 1117–1124, Jan. 2015.
[134] E. M. Redmiles et al., "A comprehensive quality evaluation of security and privacy advice on the web," in *Proc. 29th USENIX Secur. Symp. (USENIX Security)*. Berkeley, CA, USA: USENIX, 2020, pp. 89–108.
[135] R. Chen, J. Gaia, and H. R. Rao, "An examination of the effect of recent phishing encounters on phishing susceptibility," *Decis. Support Syst.*, vol. 133, Jun. 2020, Art. no. 113287.
[136] D. House and M. K. Raja, "Phishing: Message appraisal and the exploration of fear and self-confidence," *Behaviour Inf. Technol.*, vol. 39, no. 11, pp. 1204–1224, Nov. 2020.
[137] K. W. Hong, C. M. Kelley, R. Tembe, E. Murphy-Hill, and C. B. Mayhorn, "Keeping up with the Joneses: Assessing phishing susceptibility in an email task," in *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, vol. 57. Los Angeles, CA, USA: SAGE, 2013, pp. 1012–1016.
[138] S. Mondal, D. Maheshwari, N. Pai, and A. Biwalkar, "A review on detecting phishing URLs using clustering algorithms," in *Proc. Int. Conf. Adv. Comput., Commun. Control (ICAC3)*. Mumbai, India: IEEE, Dec. 2019, pp. 1–6.
[139] A. Alyahya and G. R. S. Weir, "Understanding responses to phishing in Saudi Arabia via the theory of planned behaviour," in *Proc. Nat. Comput. Colleges Conf. (NCCC)*. Taif, Saudi Arabia: IEEE, Mar. 2021, pp. 1–6.
[140] Q. Wang et al., "Adversary resistant deep neural networks with an application to malware detection," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. Halifax, NS, Canada: ACM, Aug. 2017, pp. 1145–1153.
[141] H. Aldawood and G. Skinner, "Reviewing cyber security social engineering training and awareness programs—Pitfalls and ongoing issues," *Future Internet*, vol. 11, no. 3, p. 73, Mar. 2019.
[142] U. Jensen, "Probabilistic risk analysis: Foundations and methods," *J. Amer. Stat. Assoc.*, vol. 97, no. 459, p. 925, 2002.
[143] A. M. Ness et al., "Reactions to ideological websites: The impact of emotional appeals, credibility, and pre-existing attitudes," *Comput. Hum. Behav.*, vol. 72, pp. 496–511, Jul. 2017.
[144] P. L. Gallegos-Segovia, J. F. Bravo-Torres, V. M. Larios-Rosillo, P. E. Vintimilla-Tapia, I. F. Yuquilima-Albarado, and J. D. Jara-Saltos, "Social engineering as an attack vector for ransomware," in *Proc. CHILEAN Conf. Electr., Electron. Eng., Inf. Commun. Technol. (CHILECON)*. Pucon, Chile: IEEE, Oct. 2017, pp. 1–6.
[145] N. Benias and A. P. Markopoulos, "Hacking the human: Exploiting primordial instincts," in *Proc. South-Eastern Eur. Design Autom., Comput. Eng., Comput. Netw. Soc. Media Conf. (SEEDA_CECNSM)*. Kastoria, Greece: IEEE, Sep. 2018, pp. 1–6.
[146] M. R. Arabia-Obedoza, G. Rodriguez, A. Johnston, F. Salahdine, and N. Kaabouch, "Social engineering attacks a reconnaissance synthesis analysis," in *Proc. 11th IEEE Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, Oct. 2020, pp. 843–848.
[147] V. Greavu Serban and O. Serban, "Social engineering a general approach," *Inf. Economica*, vol. 18, no. 2, pp. 5–14, Jun. 2014.
[148] C. Schürmann, L. H. Jensen, and R. M. Sigbjörnsdóttir, "Effective cybersecurity awareness training for election officials," in *Proc. Int. Joint Conf. Electron. Voting*. Bregenz, Austria: Springer, 2020, pp. 196–212.
[149] D. Airehrour, N. V. Nair, and S. Madanian, "Social engineering attacks and countermeasures in the New Zealand banking system: Advancing a user-reflective mitigation model," *Information*, vol. 9, no. 5, p. 110, May 2018.
[150] S. Buecker, M. Maes, J. J. Denissen, and M. Luhmann, "Loneliness and the big five personality traits: A meta–analysis," *Eur. J. Personality*, vol. 34, no. 1, pp. 8–28, 2020.
[151] C. Lekati, "Complexities in investigating cases of social engineering: How reverse engineering and profiling can assist in the collection of evidence," in *Proc. 11th Int. Conf. IT Secur. Incident Manage. IT Forensics (IMF)*. Hamburg, Germany: IEEE, May 2018, pp. 107–109.
[152] J. Deutrom, V. Katos, and R. Ali, "Loneliness, life satisfaction, problematic Internet use and security behaviours: Re-examining the relationships when working from home during COVID-19," *Behaviour Inf. Technol.*, vol. 41, no. 14, pp. 3161–3175, Oct. 2022.
[153] D. Conway, R. Taib, M. Harris, K. Yu, S. Berkovsky, and F. Chen, "A qualitative investigation of bank employee experiences of information security and phishing," in *Proc. 13th Symp. Usable Privacy Secur. (SOUPS)*. Berkeley, CA, USA: USENIX, 2017, pp. 115–129.
[154] B. Brinton Anderson, A. Vance, C. B. Kirwan, D. Eargle, and J. L. Jenkins, "How users perceive and respond to security messages: A NeuroIS research agenda and empirical study," *Eur. J. Inf. Syst.*, vol. 25, no. 4, pp. 364–390, Jul. 2016.
[155] J. W. Stoughton, L. F. Thompson, and A. W. Meade, "Big five personality traits reflected in job applicants' social media postings," *Cyberpsychol., Behav., Social Netw.*, vol. 16, no. 11, pp. 800–805, Nov. 2013.
[156] W. Lu, S. Xu, and X. Yi, "Optimizing active cyber defense," in *Proc. Int. Conf. Decis. Game Theory Secur.* Cham, Switzerland: Springer, 2013, pp. 206–225.
[157] A. Aleroud, E. Abu-Shanab, A. Al-Aiad, and Y. Alshboul, "An examination of susceptibility to spear phishing cyber attacks in non-english speaking communities," *J. Inf. Secur. Appl.*, vol. 55, Dec. 2020, Art. no. 102614.
[158] P. K. Yeng, M. A. Fauzi, B. Yang, and P. Nimbe, "Investigation into phishing risk behaviour among healthcare staff," *Information*, vol. 13, no. 8, p. 392, Aug. 2022.
[159] M. L. Jensen, A. Durcikova, and R. T. Wright, "Using susceptibility claims to motivate behaviour change in IT security," *Eur. J. Inf. Syst.*, vol. 30, no. 1, pp. 27–45, Jan. 2021.
[160] T. Grassegger and D. Nedbal, "The role of employees' information security awareness on the intention to resist social engineering," *Proc. Comput. Sci.*, vol. 181, pp. 59–66, Jan. 2021.
[161] H. Shahbaznezhad, F. Kolini, and M. Rashidirad, "Employees' behavior in phishing attacks: What individual, organizational, and technological factors matter?" *J. Comput. Inf. Syst.*, vol. 61, no. 6, pp. 1–12, Sep. 2020.
[162] G. Diksha and J. A. Kumar, "Mobile phishing attacks and defence mechanisms: State of art and open research challenges," *Comput. Secur.*, vol. 73, pp. 519–544, Mar. 2018.
[163] M. Alohali, N. Clarke, F. Li, and S. Furnell, "Identifying and predicting the factors affecting end-users' risk-taking behavior," *Inf. Comput. Secur.*, vol. 26, no. 3, pp. 306–326, Jul. 2018.
[164] R. T. Wright, M. L. Jensen, J. B. Thatcher, M. Dinger, and K. Marett, "Research note—Influence techniques in phishing attacks: An examination of vulnerability and resistance," *Inf. Syst. Res.*, vol. 25, no. 2, pp. 385–400, Jun. 2014.
[165] M. Junger, V. Wang, and M. Schlömer, "Fraud against businesses both online and offline: Crime scripts, business characteristics, efforts, and benefits," *Crime Sci.*, vol. 9, no. 1, pp. 1–15, Dec. 2020.
[166] R. Dhamija, J. D. Tygar, and M. Hearst, "Why phishing works," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.* Montreal, QC, Canada: ACM, 2006, pp. 581–590.
[167] A. Vishwanath, T. Herath, R. Chen, J. Wang, and H. R. Rao, "Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model," *Decis. Support Syst.*, vol. 51, no. 3, pp. 576–586, Jun. 2011.
[168] R. M. Rodriguez, A. Atyabi, and Shouhuai, "Social engineering attacks and defenses in the physical world vs. cyberspace: A contrast study," 2022, *arXiv:2203.04813*.
[169] C. Herley, "Why do Nigerian scammers say they are from Nigeria?" in *Proc. Workshop Econ. Inf. Secur.*, Berlin, Germany, 2012, pp. 1–14.
[170] T. Nelms, R. Perdisci, M. Antonakakis, and M. Ahamad, "Towards measuring and mitigating social engineering software download attacks," in *Proc. 25th USENIX Security Symp. (USENIX Security)*. Austin, TX, USA: USENIX Association, 2016, pp. 773–789.
[171] J. B. Hirsh, S. K. Kang, and G. V. Bodenhausen, "Personalized persuasion: Tailoring persuasive appeals to recipients' personality traits," *Psychol. Sci.*, vol. 23, no. 6, pp. 578–581, Jun. 2012.
[172] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer, "Social phishing," *Commun. ACM*, vol. 50, no. 10, pp. 94–100, 2007.
[173] J. L. Freedman and S. C. Fraser, "Compliance without pressure: The foot-in-the-door technique," *J. Personality Social Psychol.*, vol. 4, no. 2, pp. 195–202, 1966.
[174] V. Tiwari, "Analysis and detection of fake profile over social network," in *Proc. Int. Conf. Comput., Commun. Autom. (ICCCA)*, May 2017, pp. 175–179.
[175] L. Allodi, T. Chotza, E. Panina, and N. Zannone, "The need for new antiphishing measures against spear-phishing attacks," *IEEE Secur. Privacy*, vol. 18, no. 2, pp. 23–34, Mar. 2020.
[176] D. J. McAllister, "Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations," *Acad. Manage. J.*, vol. 38, no. 1, pp. 24–59, Feb. 1995.
[177] M. Cui, "How does the decoy effect affect decision-making and how we can prevent it?" in *Proc. 7th Int. Conf. Financial Innov. Econ. Develop. (ICFIED)*. Amsterdam, The Netherlands: Atlantis Press, 2022, pp. 1753–1756.
[178] T. Seitz, E. von Zezschwitz, S. Meitner, and H. Hussmann, "Influencing self-selected passwords through suggestions and the decoy effect," in *Proc. 1st Eur. Workshop Usable Secur.*, vol. 2, 2016, pp. 1–2.
[179] M. Junger, L. Montoya, and F.-J. Overink, "Priming and warnings are not effective to prevent social engineering attacks," *Comput. Hum. Behav.*, vol. 66, pp. 75–87, Jan. 2017.
[180] O. Gillath and G. Karantzas, "Attachment security priming: A systematic review," *Current Opinion Psychol.*, vol. 25, pp. 86–95, Feb. 2019.
[181] (Jun. 12, 2019). *5 Phishing Emails That Led to Real-World Data Breaches*. Accessed: Dec. 8, 2023. [Online]. Available: https://resources.infosecinstitute.com/topics/phishing/5-phishing-emails-that-led-to-real-world-data-breaches/
[182] M. M. Ali and N. F. M. Zaharon, "Phishing—A cyber fraud: The types, implications and governance," *Int. J. Educ. Reform*, vol. 33, no. 1, pp. 101–121, 2022.
[183] R. Abid, M. Rizwan, P. Veselý, A. Basharat, U. Tariq, and A. R. Javed, "Social networking security during COVID-19: A systematic literature review," *Wireless Commun. Mobile Comput.*, vol. 2022, pp. 1–21, Apr. 2022.
[184] L. Huang and Q. Zhu, "A dynamic games approach to proactive defense strategies against advanced persistent threats in cyber-physical systems," *Comput. Secur.*, vol. 89, Feb. 2020, Art. no. 101660.

[185] A. Bhardwaj, V. Sapra, A. Kumar, N. Kumar, and S. Arthi, "Why is phishing still successful?" *Comput. Fraud Secur.*, vol. 2020, no. 9, pp. 15–19, Jan. 2020.

[186] CISA. (2023). *China Cyber Threat Overview and Advisories: CISA*. [Online]. Available: https://www.cisa.gov/china

[187] A. Chitrey, D. Singh, and V. Singh, "A comprehensive study of social engineering based attacks in India to develop a conceptual model," *Int. J. Inf. Netw. Secur.*, vol. 1, no. 2, p. 45, Jun. 2012.

[188] G. Ho, "Detecting and characterizing lateral phishing at scale," in *Proc. 28th USENIX Secur. Symp. (USENIX Security)*. Berkeley, CA, USA: USENIX, 2019, pp. 1273–1290.

[189] M. P. Steves, K. K. Greene, and M. F. Theofanos, "A phish scale: Rating human phishing message detection difficulty," in *Proc. Workshop Usable Secur. (USEC)*, 2019, pp. 1–14.

[190] A. Vishwanath, "Getting phished on social media," *Decis. Support Syst.*, vol. 103, pp. 70–81, Nov. 2017.

[191] G. Ho, A. Sharma, M. Javed, V. Paxson, and D. Wagner, "Detecting credential spearphishing in enterprise settings," in *Proc. 26th USENIX Secur. Symp. (USENIX Security)*. Berkeley, CA, USA: USENIX, 2017, pp. 469–485.

[192] V. Bhavsar, A. Kadlak, and S. Sharma, "Study on phishing attacks," *Int. J. Comput. Appl.*, vol. 182, pp. 27–29, Dec. 2018.

[193] M. N. Alam, D. Sarma, F. F. Lima, I. Saha, Rubaiath-E-Ulfath, and S. Hossain, "Phishing attacks detection using machine learning approach," in *Proc. 3rd Int. Conf. Smart Syst. Inventive Technol. (ICSSIT)*. Tirunelveli, India: IEEE, Aug. 2020, pp. 1173–1179.

[194] S. P. Prem and B. I. Reddy, "Phishing and anti-phishing techniques," *Int. Res. J. Eng. Technol.*, vol. 6, no. 7, pp. 1446–1452, 2019.

[195] G. Burch, A. Taylor, and C. Yeung, "Wire transfer email fraud and what to do about it," *Intellectual Property Technol. Law J.*, vol. 27, no. 1, p. 13, 2015.

[196] R. Chaganti, B. Bhushan, A. Nayyar, and A. Mourade, "Recent trends in social engineering scams and case study of gift card scam," 2021, *arXiv:2110.06487*.

[197] A. Cidon, L. Gavish, I. Bleier, N. Korshun, M. Schweighauser, and A. Tsitkin, "High precision detection of business email compromise," in *Proc. 28th USENIX Secur. Symp. (USENIX Security)*. Berkeley, CA, USA: USENIX, 2019, pp. 1291–1307.

[198] S. Venkatesha, K. R. Reddy, and B. R. Chandavarkar, "Social engineering attacks during the COVID-19 pandemic," *Social Netw. Comput. Sci.*, vol. 2, no. 2, pp. 1–9, Apr. 2021.

[199] N. Miramirkhani, O. Starov, and N. Nikiforakis, "Dial one for scam: A large-scale analysis of technical support scams," 2016, *arXiv:1607.06891*.

[200] O. Or-Meir, N. Nissim, Y. Elovici, and L. Rokach, "Dynamic malware analysis in the modern era—A state of the art survey," *ACM Comput. Surv.*, vol. 52, no. 5, pp. 1–48, Sep. 2020.

[201] B. Srinivasan et al., "Exposing search and advertisement abuse tactics and infrastructure of technical support scammers," in *Proc. World Wide Web Conf.*, 2018, pp. 319–328.

[202] J. Meinert, M. Mirbabaie, S. Dungs, and A. Aker, "Is it really fake? Towards an understanding of fake news in social media communication," in *Proc. Int. Conf. Social Comput. Social Media*. Springer, 2018, pp. 484–497.

[203] D. López-Sánchez, J. R. Herrero, A. G. Arrieta, and J. M. Corchado, "Hybridizing metric learning and case-based reasoning for adaptable clickbait detection," *Appl. Intell.*, vol. 48, no. 9, pp. 2967–2982, Sep. 2018.

[204] P. Vadrevu and R. Perdisci, "What you see is NOT what you get: Discovering and tracking social engineering attack campaigns," in *Proc. Internet Meas. Conf.*, 2019, pp. 308–321.

[205] K. Subramani, X. Yuan, O. Setayeshfar, P. Vadrevu, K. H. Lee, and R. Perdisci, "When push comes to ads: Measuring the rise of (malicious) push advertising," in *Proc. ACM Internet Meas. Conf.*, Oct. 2020, pp. 724–737.

[206] N. Provos, D. McNamee, P. Mavrommatis, K. Wang, and N. Modadugu, "The ghost in the browser analysis of web-based malware," in *Proc. 1st Workshop Hot Topics Understand. Botnets (HotBots)*, 2007.

[207] D. Irani, M. Balduzzi, D. Balzarotti, E. Kirda, and C. Pu, "Reverse social engineering attacks in online social networks," in *Proc. Int. Conf. Detection Intrusions Malware, Vulnerability Assessment*. Springer, 2011, pp. 55–74.

[208] M. Z. Rafique, T. Van Goethem, W. Joosen, C. Huygens, and N. Nikiforakis, "It's free for a reason: Exploring the ecosystem of free live streaming services," in *Proc. Netw. Distrib. Syst. Secur. Symp.* Reston, VA, USA: Internet Society, 2016, pp. 1–15.

[209] E. Velasquez. (Jun. 2018). *What is Angler Phishing and How Can You Avoid It?* [Online]. Available: https://www.experian.com/blogs/ask-experian/what-is-angler-phishing-and-how-can-you-avoid-it/

[210] L. O'Hagan, "Angler phishing: Criminality in social media," in *Proc. 5th Eur. Conf. Social Media (ECSM)*, 2018, p. 190.

[211] M. A. Ivanov, B. V. Kliuchnikova, I. V. Chugunkov, and A. M. Plaksina, "Phishing attacks and protection against them," in *Proc. IEEE Conf. Russian Young Res. Electr. Electron. Eng. (ElConRus)*, Jan. 2021, pp. 425–428.

[212] M. Steffens, C. Rossow, M. Johns, and B. Stock, "Don't trust the locals: Investigating the prevalence of persistent client-side cross-site scripting in the wild," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, 2019.

[213] F. Kanei, D. Chiba, K. Hato, K. Yoshioka, T. Matsumoto, and M. Akiyama, "Detecting and understanding online advertising fraud in the wild," *IEICE Trans. Inf. Syst.*, vol. 103, no. 7, pp. 1512–1523, 2020.

[214] R. O. Oveh and G. O. Aziken, "Mitigating social engineering attack: A focus on the weak human link," in *Proc. 5th Inf. Technol. Educ. Develop. (ITED)*, Nov. 2022, pp. 1–4.

[215] R. F. Abu Hweidi and D. Eleyan, "Social engineering attack concepts, frameworks, and awareness: A systematic literature review," *Int. J. Comput. Digit. Syst.*, vol. 13, no. 1, pp. 691–700, Apr. 2023.

[216] T. Copado. (Aug. 2021). *12 Types of Social Engineering Attacks to Look Out For*. [Online]. Available: https://www.copado.com/devops-hub/blog/12-types-of-social-engineering-attacks-to-look-out-for

[217] M. Simmons and J. S. Lee, "Catfishing: A look into online dating and impersonation," in *Proc. Int. Conf. Human-Comput. Interact.* Oldenburg, Germany: Springer, 2020, pp. 349–358.

[218] C. Lauder and E. March, "Catching the catfish: Exploring gender and the dark tetrad of personality as predictors of catfishing perpetration," *Comput. Hum. Behav.*, vol. 140, Mar. 2023, Art. no. 107599.

[219] L. Malisa, K. Kostiainen, and S. Capkun, "Detecting mobile application spoofing attacks by leveraging user visual similarity perception," in *Proc. 7th ACM Conf. Data Appl. Secur. Privacy*. New York, NY, USA: ACM, Mar. 2017, pp. 289–300.

[220] Y. Hu, G. Xu, B. Zhang, K. Lai, G. Xu, and M. Zhang, "Robust app clone detection based on similarity of UI structure," *IEEE Access*, vol. 8, pp. 77142–77155, 2020.

[221] S. Calzavara, S. Roth, A. Rabitti, M. Backes, and B. Stock, "A tale of two headers: A formal analysis of inconsistent click-jacking protection on the web," in *Proc. 29th USENIX Secur. Symp. (USENIX Security)*, 2020, pp. 683–697.

[222] A. K. Jain and B. B. Gupta, "Rule-based framework for detection of Smishing messages in mobile environment," *Proc. Comput. Sci.*, vol. 125, pp. 617–623, Jan. 2018.

[223] T. Xu et al., "Deep entity classification: Abusive account detection for online social networks," in *Proc. 30th USENIX Secur. Symp. (USENIX Security)*, 2021.

[224] R. Verma, N. Shashidhar, and N. Hossain, "Detecting phishing emails the natural language way," in *Proc. 17th Eur. Symp. Res. Comput. Secur. (ESORICS)*. Springer, 2012, pp. 824–841.

[225] D. He et al., "An effective double-layer detection system against social engineering attacks," *IEEE Netw.*, vol. 36, no. 6, pp. 92–98, Nov. 2022.

[226] Z. Ling, H. Feng, X. Ding, X. Wang, C. Gao, and P. Yang, "Spear phishing email detection with multiple reputation features and sample enhancement," in *Proc. 4th Int. Conf. Sci. Cyber Security (SciSec)*, Matsue, Japan. Springer, 2022, pp. 522–538.

[227] N. Tsinganos, P. Fouliras, and I. Mavridis, "Applying BERT for early-stage recognition of persistence in chat-based social engineering attacks," *Appl. Sci.*, vol. 12, no. 23, p. 12353, Dec. 2022.

[228] J. Lee, F. Tang, P. Ye, F. Abbasi, P. Hay, and D. M. Divakaran, "D-fence: A flexible, efficient, and comprehensive phishing email detection system," in *Proc. IEEE Eur. Symp. Secur. Privacy (EuroS&P)*, Sep. 2021, pp. 578–597.

[229] Y. Lin et al., "Phishpedia: A hybrid deep learning based approach to visually identify phishing webpages," in *Proc. 30th Usenix Secur. Symp.*, 2021.

[230] S. Abdelnabi, K. Krombholz, and M. Fritz, "VisualPhishNet: Zero-day phishing website detection by visual similarity," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.* New York, NY, USA: Association for Computing Machinery, Oct. 2020, pp. 1681–1698, doi: 10.1145/3372297.3417233.

[231] L. Goeke, A. Quintanar, K. Beckers, and S. Pape, "PROTECT—An easy configurable serious game to train employees against social engineering attacks," in *Computer Security*. Springer, 2020, pp. 156–171.

[232] D. Aladawy, K. Beckers, and S. Pape, "Persuaded: Fighting social engineering attacks with a serious game," in *Proc. Int. Conf. Trust Privacy Digital Bus.*, Bratislava, Slovakia. Cham, Switzerland: Springer, 2018, pp. 103–118.

[233] J. Mao, P. Li, K. Li, T. Wei, and Z. Liang, "BaitAlarm: Detecting phishing sites using similarity in fundamental visual features," in *Proc. 5th Int. Conf. Intell. Netw. Collaborative Syst.*, Sep. 2013, pp. 790–795.

[234] C. C. L. Tan, K. L. Chiew, K. S. C. Yong, Y. Sebastian, J. C. M. Than, and W. K. Tiong, "Hybrid phishing detection using joint visual and textual identity," *Expert Syst. Appl.*, vol. 220, Jun. 2023, Art. no. 119723.

[235] A. Nakamura and F. Dobashit, "Proactive phishing sites detection," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*. Thessaloniki, Greece: IEEE, Oct. 2019, pp. 443–448.

[236] A. Kharraz, W. Robertson, and E. Kirda, "Surveylance: Automatically detecting online survey scams," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2018, pp. 70–86.

[237] G. Jethava and U. P. Rao, "A novel defense mechanism to protect users from profile cloning attack on online social networks (OSNs)," *Peer-to-Peer Netw. Appl.*, vol. 15, no. 5, pp. 2253–2269, Sep. 2022.

[238] N. Tsinganos, I. Mavridis, and D. Gritzalis, "Utilizing convolutional neural networks and word embeddings for early-stage recognition of persuasion in chat-based social engineering attacks," *IEEE Access*, vol. 10, pp. 108517–108529, 2022.

[239] M. Lansley, F. Mouton, S. Kapetanakis, and N. Polatidis, "SEADer++: Social engineering attack detection in online environments using machine learning," *J. Inf. Telecommun.*, vol. 4, no. 3, pp. 346–362, Jul. 2020.

[240] D. Kahneman, *Thinking, Fast Slow*. New York, NY, USA: Macmillan, 2011.

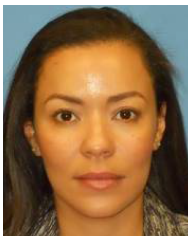
[241] W. De Neys and G. Pennycook, "Logic, fast and

slow: Advances in dual-process theorizing," *Current Directions Psychol. Sci.*, vol. 28, no. 5, pp. 503–509, Oct. 2019.
[242] I. Del Pozo, M. Iturralde, and F. Restrepo, "Social engineering: Application of psychology to information security," in *Proc. 6th Int. Conf. Future Internet Things Cloud Workshops (FiCloudW)*. Washington, DC, USA: IEEE Computer Society, Aug. 2018, pp. 108–114.
[243] J. A. Cummings and L. Sanders, *Introduction to Psychology*. University of Saskatchewan Open Press, 2019.
[244] C. C. Williams, M. Kappen, C. D. Hassall, B. Wright, and O. E. Krigolson, "Thinking theta and alpha: Mechanisms of intuitive and analytical reasoning," *NeuroImage*, vol. 189, pp. 574–580, Apr. 2019.
[245] G. Pennycook, J. A. Fugelsang, and D. J. Koehler, "What makes us think? A three-stage dual-process model of analytic engagement," *Cogn. Psychol.*, vol. 80, pp. 34–72, Aug. 2015.
[246] H. Lin, G. Pennycook, and D. G. Rand, "Thinking more or thinking differently? Using drift-diffusion modeling to illuminate why accuracy prompts decrease misinformation sharing," *Cognition*, vol. 230, Jan. 2023, Art. no. 105312.
[247] G. Booch et al., "Thinking fast and slow in AI," in *Proc. AAAI Conf. Artif. Intell.*, 2020.
[248] A. Abbasi, F. M. Zahedi, and Y. Chen, "Phishing susceptibility: The good, the bad, and the ugly," in *Proc. IEEE Conf. Intell. Secur. Informat. (ISI)*. Tucson, AZ, USA: IEEE, Sep. 2016, pp. 169–174.
[249] P. Kumaraguru, A. Acquisti, and L. F. Cranor, "Trust modelling for online transactions: A phishing scenario," in *Proc. Int. Conf. Privacy, Secur. Trust, Bridge Gap Between PST Technol. Bus. Services*. Markham, ON, Canada: ACM, Oct. 2006, p. 11.
[250] R. B. Cialdini and M. R. Trost, "Social influence: Social norms, conformity and compliance," in *The Handbook of Social Psychology*. New York, NY, USA: McGraw-Hill, 1998, p. 151.
[251] A. Kirmani and R. Zhu, "Vigilant against manipulation: The effect of regulatory focus on the use of persuasion knowledge," *J. Marketing Res.*, vol. 44, no. 4, pp. 688–701, Nov. 2007.
[252] A. H. Maslow, "A theory of human motivation," *Psychol. Rev.*, vol. 50, no. 4, p. 370, 1943.
[253] A. E. Howe, I. Ray, M. Roberts, M. Urbanska, and Z. Byrne, "The psychology of security for the home computer user," in *Proc. IEEE Symp. Secur. Privacy*. Washington, DC, USA: IEEE Computer Society, May 2012, pp. 209–223, doi: 10.1109/SP2012.23.
[254] Y. J. Kim, R. Kishore, and G. L. Sanders, "From DQ to EQ: Understanding data quality in the context of e-business systems," *Commun. ACM*, vol. 48, no. 10, pp. 75–81, Oct. 2005.
[255] M. S. Wogalter, "Communication-human information processing (C-HIP) model," in *Forensic Human Factors and Ergonomics*. Boca Raton, FL, USA: CRC Press, 2018, pp. 33–49.
[256] R. Heartfield, G. Loukas, and D. Gan, "You are probably not the weakest link: Towards practical prediction of susceptibility to semantic social engineering attacks," *IEEE Access*, vol. 4, pp. 6910–6928, 2016.
[257] A. Alturki, N. Alshwihi, and A. Algarni, "Factors influencing players' susceptibility to social engineering in social gaming networks," *IEEE Access*, vol. 8, pp. 97383–97391, 2020.
[258] S. Xu, "Cybersecurity dynamics," in *Proc. Symp. Bootcamp Sci. Secur.* New York, NY, USA: ACM, Apr. 2014, pp. 1–2.
[259] S. Xu, "Cybersecurity dynamics: A foundation for the science of cybersecurity," in *Proactive and Dynamic Network Defense*. Springer, 2019, pp. 1–31.
[260] S. Xu, "The cybersecurity dynamics way of thinking and landscape (invited paper)," in *Proc. 7th ACM Workshop Moving Target Defense*. New York, NY, USA: ACM, Nov. 2020, pp. 69–80.
[261] S. Xu, "SARR: A cybersecurity metrics and quantification framework (keynote)," in *Proc. 3rd Int. Conf. Sci. Cyber Secur. (SciSec)*, Shanghai, China, in Lecture Notes in Computer Science, vol. 13005. Cham, Switzerland: Springer, 2021, pp. 3–17.
[262] M. Pendleton, R. Garcia-Lebron, J.-H. Cho, and S. Xu, "A survey on systems security metrics," *ACM Comput. Surv.*, vol. 49, no. 4, pp. 1–35, Dec. 2017.
[263] J. Cho, S. Xu, P. Hurley, M. Mackay, T. Benjamin, and M. Beaumont, "STRAM: Measuring the trustworthiness of computer-based systems," *ACM Comput. Surv.*, vol. 51, no. 6, pp. 128:1–128:47, 2019.
[264] X. Li, P. Parker, and S. Xu, "A stochastic model for quantitative security analyses of networked systems," *IEEE Trans. Dependable Secure Comput.*, vol. 8, no. 1, pp. 28–43, Jan. 2011.
[265] S. Xu, W. Lu, and L. Xu, "Push- and pull-based epidemic spreading in networks: Thresholds and deeper insights," *ACM Trans. Auto. Adapt. Syst.*, vol. 7, no. 3, pp. 1–26, Sep. 2012.
[266] S. Xu, W. Lu, and Z. Zhan, "A stochastic model of multivirus dynamics," *IEEE Trans. Dependable Secure Comput.*, vol. 9, no. 1, pp. 30–45, Jan. 2012.
[267] R. Zheng, W. Lu, and S. Xu, "Preventive and reactive cyber defense dynamics is globally stable," *IEEE Trans. Netw. Sci. Eng.*, vol. 5, no. 2, pp. 156–170, Apr. 2018.
[268] Z. Lin, W. Lu, and S. Xu, "Unified preventive and reactive cyber defense dynamics is still globally convergent," *IEEE/ACM Trans. Netw.*, vol. 27, no. 3, pp. 1098–1111, Jun. 2019.
[269] Y. Han, W. Lu, and S. Xu, "Preventive and reactive cyber defense dynamics with ergodic time-dependent parameters is globally attractive," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 3, pp. 2517–2532, Jul. 2021.
[270] S. Xu, W. Lu, L. Xu, and Z. Zhan, "Adaptive epidemic dynamics in networks: Thresholds and control," *ACM Trans. Auto. Adapt. Syst.*, vol. 8, no. 4, pp. 1–19, Jan. 2014.
[271] G. Da, M. Xu, and S. Xu, "A new approach to modeling and analyzing security of networked systems," in *Proc. HotSoS*. New York, NY, USA: ACM, Apr. 2014, pp. 6:1–6:12.
[272] Y. Han, W. Lu, and S. Xu, "Characterizing the power of moving target defense via cyber epidemic dynamics," in *Proc. HotSoS*. New York, NY, USA: ACM, Apr. 2014, pp. 1–12.
[273] S. Xu, W. Lu, and H. Li, "A stochastic model of active cyber defense dynamics," *Internet Math.*, vol. 11, no. 1, pp. 23–61, Jan. 2015.
[274] R. Zheng, W. Lu, and S. Xu, "Active cyber defense dynamics exhibiting rich phenomena," in *Proc. Symp. Bootcamp Sci. Secur.* New York, NY, USA: ACM, Apr. 2015, pp. 1–12.
[275] M. Xu and S. Xu, "An extended stochastic model for quantitative security analysis of networked systems," *Internet Math.*, vol. 8, no. 3, pp. 288–320, Aug. 2012.
[276] M. Xu, G. Da, and S. Xu, "Cyber epidemic models with dependences," *Internet Math.*, vol. 11, no. 1, pp. 62–92, Jan. 2015.
[277] H. Chen, J. Cho, and S. Xu, "Quantifying the security effectiveness of firewalls and DMZs," in *Proc. HoTSoS*. New York, NY, USA: ACM, 2018, pp. 9:1–9:11.
[278] H. Chen, H. Cam, and S. Xu, "Quantifying cybersecurity effectiveness of dynamic network diversity," *IEEE Trans. Dependable Secure Comput.*, vol. 19, no. 6, pp. 3804–3821, Nov. 2022.

ABOUT THE AUTHORS

Theodore Tangie Longtchi received the B.S. degree in computer science and software engineering with a focus on information assurance and cybersecurity and the M.S. degree in cybersecurity and leadership from the University of Washington, Seattle, WA, USA, in March 2016 and August 2017, respectively. He is currently working toward the Ph.D. degree at the University of Colorado Colorado Springs, Colorado Springs, CO, USA.

At the University of Colorado Colorado Springs, his research interests are in social engineering, cybersecurity, cognitive psychology, and the different domains of computer and information security. He also holds the CompTIA Security+ and the Certified Information Systems Security Professional (CISSP) certifications.



Rosana Montañez Rodriguez received the bachelor's degree in applied science from the University of Puerto Rico (Rio Piedras Campus), San Juan, Puerto Rico, in 2000, the master's degree in security engineering from Southern Methodist University, Dallas, TX, USA, in 2012, and the bachelor's degree in computer science from the University of Maryland (University Campus), College Park, MD, USA, in 2013. She is currently working toward the Ph.D. degree at The University of Texas at San Antonio, San Antonio, TX, USA.

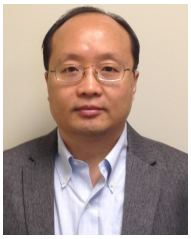
She has over 20 years of professional experience in IT operations and engineering, software engineering, and cybersecurity and has been a certified Certified Information Systems Security Professional (CISSP) since 2014. She is currently a Cybersecurity Engineer with The MITRE Corporation, San Antonio. Her research explores the connection between human factors, cognitive psychology, and cybersecurity. Her interests include cognition and security performance, computer-mediated information interpretation, and knowledge networks.

Laith Al-Shawaf is currently an Associate Professor with the Department of Psychology, University of Colorado Colorado Springs, Colorado Springs, CO, USA, and a Visiting Fellow with the Institute for Advanced Study in Toulouse (IAST), Toulouse, France. Before moving to the United States, he held an academic position in Turkey and was a Visiting Fellow with the Institute for Advanced Study in Berlin (Wiko), Berlin, Germany. His empirical research centers on human emotions, personality and individual differences, and cognitive biases. He often conducts research cross-culturally. His popular science work has been translated into several languages.

Dr. Al-Shawaf has won awards for both teaching and research. He is the Primary Editor of *The Oxford Handbook of Evolution and the Emotions*.

Adham Atyabi (Member, IEEE) received the Ph.D. degree from Flinders University, Adelaide, SA, Australia, in 2013.

He is currently an Assistant Professor with the University of Colorado Colorado Springs, Colorado Springs, CO, USA. His research interests include autism biomarker discovery, cognitive and computational neuroscience, computational intelligence, image and signal processing, and swarm and cognitive robotics.



Shouhuai Xu (Senior Member, IEEE) received the Ph.D. degree in computer science from Fudan University, Shanghai, China, in 2000.

He is currently the Gallogly Chair Professor with the Department of Computer Science, University of Colorado Colorado Springs (UCCS), Colorado Springs, CO, USA. He pioneered the cybersecurity dynamics approach as the foundation for the emerging science of cybersecurity, with three pillars: first-principle cybersecurity modeling and analysis; cybersecurity data analytics; and cybersecurity metrics.

Dr. Xu coinitiated the International Conference on Science of Cyber Security and is serving as its Steering Committee Chair. He is/was an Associate Editor of IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING (IEEE TDSC), IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY (IEEE T-IFS), and IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING (IEEE TNSE).