Sensitive Data? Now That's a Catch! the Psychology of Phishing

Sensitive Data? Now That's a Catch! The Psychology of Phishing

Jason Sibrian

Advisor: Professor James Mickens

Harvard University
John A. Paulson School of Engineering and Applied Sciences
April 3, 2020

Abstract

Humans are generally well adapted to handle a majority of threats that they are faced with. The issue is that behaviors that are adaptive and useful for most human interactions do not always transfer well to the digital world. In fact, our brains seem to be hardwired to fall prey to phishing. But why is that?

Phishing takes advantage of the manners in which we make decisions, how we handle our emotions, and the ways we can subconsciously be persuaded. No one is immune from these attacks, and given that our brains likely will not rewire anytime soon, phishing looks like it is around to stay. However, humans are not entirely helpless when it comes to phishing. This thesis proposes behavioral, technological, and societal mitigations that can help to decrease the susceptibility to phishing and the damage it causes when it is successful.

Acknowledgements

There is a long list of people who I would like to thank, but I will try to keep this short.

I am deeply grateful to my advisor, James Mickens, for the guidance he has provided on this journey. His course on systems security reinforced my interest on the topic and led me to decide what my thesis was to be written on. I am also very grateful to all the professors and mentors I have had throughout my years at Harvard.

I would like to thank the Mind, Brain, and Behavior Initiative for so aptly capturing an interdisciplinary study plan, allowing me to study exactly what I was interested in with little compromise.

I would like to thank the friends I have made in my time at Harvard. They have made my experience here enjoyable and have given me memories and friendships I will cherish forever. I would specifically like to thank my friend Abigail Ory in the context of this thesis, who took upon herself the momentous task of editing my (at some points less-than-stellar) grammar.

Lastly, I would like to thank my parents, Bianca and Luis. It is because of their support that I am who I am and that I am where I am today. They raised me well and gave me the skills necessary to succeed. I am–and will forever be–grateful to them. I love you Mom and Dad!

Contents

List of Figures

Chapter 1

Introduction

The concept of phishing was first mentioned on January 2nd, 1996 by a newsgroup called AOHell [1]-[2]. This was fitting as one of the first attacks of this sort was carried out by hackers stealing America Online accounts and passwords. As defined by Elledge, phishing is the fraudulent practice of sending emails purporting to be from reputable people or companies in order to induce individuals to reveal personal information, such as passwords and credit card numbers [2]. One of the most commonly known scams of this kind is the Nigerian 419 scam [3]. What makes phishing so successful–and dangerous–is that these attackers do not just prey on our emails. They prey on us, the user. When it comes to phishing, the weakest link in the security system is ourselves.

Phishing uses an analogy to the sport of fishing: the email (or other attack vector, such as texts and phone calls) is the bait. The users and their data are the fish, and the attacker the fisherman. Like real life fishing, phishing is not perfect and not everybody falls for the bait. But a few people do, and like for the fisherman's catch, it never goes well for the victims of these attacks. Retruster's

"2019 Phishing Statistics and Email Fraud Statistics" paints an alarming portrait of the impact such security breaches had on companies in 2019 [4]:

- The average financial cost of a data breach is $3.86 million
- Phishing accounts for 90% of data breaches
- Phishing attempts have grown 65% in the last year
- Around 1.5 million new phishing sites are created each month
- 76% of businesses reported being a victim of a phishing attack in the last year

Retruster's article also states that 30% of the phishing messages, including the ones that cause such problems for these corporations, are opened by targeted users. And once a user becomes a victim, there is a chance of it becoming a pattern. 15% of people successfully phished will be targeted at least one more time within the year [4].

Phishing can be deadly both personally and professionally. One misclick, lapse of judgement, or second of inattention can lead to life altering consequences. Phishing can result in identity theft, private client data leakage, or sensitive government information being compromised.

Phishing is particularly dangerous because it preys on human emotions and mental shortcuts, and it does so by using deception. Effectively, phishing is a high tech con and these "phishers" are the con artists.

Chapter 2

Background

To understand phishing, we must understand the broader category of cyber-attacks it falls under: hacking. Hacking is when an attacker gains unauthorized access to a computer system. In this paper the term “hacker” refers to those who have malicious intent when they gain this unauthorized access, but this is not always the case. The methods used by hackers can vary greatly, from direct attempts to exploit flaws in a systems security and programming to attacks that use the end users as vectors to circumvent security altogether. An attacker's end goal is also variable. Some like creating chaos, others search for monetary gain, and some hack in order to uphold their principles (ex. attacking large companies whose values they disagree with). Hacking can be used to compromise entire systems or just the account(s) of a single person. Whatever the goal, hacking is very dangerous, but what is more dangerous is that hacking does not stop when you move offline. Hackers are not limited to attacking from behind their computer screens. By using a skill set known as “social engineering,” attackers can interact with targets in the physical world and create interactions that facilitate these targets getting hacked online.

2.1 Traditional Hacking

Traditional hacking is a term that will be used for clarity in the following section in order to distinguish from social engineering. Traditional hacking refers to attacks that focus more on exploiting the technology rather than on exploiting the people that use it.

2.1.1 Malware

Deriving from the term “malicious software,” malware is any program that is designed to cause damage to a computer, server, or network [11]. Somewhat of a catch-all, malware includes worms, trojan horses, and viruses. In essence, if it can cause some sort of damage and can be run on a computer, it is probably malware. Malware is meant to be destructive, but what kind of damage is it capable of? Physical damage is one type: loss of data, damage to hardware, or corruption of the hardware. However, there many other types of damage malware can create: financial damage due to identity theft, social damage due to malicious social media account takeover, and emotional damage such as PTSD from being spied on through your computer’s camera are just a few. So, while not all destructive programs are created to make a computer implode, it is malware nonetheless.

2.1.1.1 Example: Key-Logging

One specific type of malware is a keylogger. These are programs designed to live on a computer and record a user's keystrokes. In doing so, they are eventually able to compromise a user's login credentials. This can be done by using the keylogger to check the web address or destination address entered prior

to the username and password that were collected. This may also be done more easily if a keylogger is part of a larger exploit like a RAT (Remote Access Tool) that may include screen recording.

2.2 Social Engineering

While traditional hacking typically focuses less on the user and more on the software, social engineering focuses less on the system and instead has users act as unwitting accomplices. Thus, social engineering can be looked at as the opposite side of the hacking coin. Typically, social engineering, as its name implies, takes advantage of the social nature of human beings to achieve its end goal. In this variant of hacking, social psychology is the key lever that forces the users of a system to compromise it for the attacker. Of course, these types of attacks can use code and technical exploits in conjunction with social psychology; however, they often do not need to be as technical as a traditional hack because some of the security measures that would typically need to be bypassed are already taken care of by piggybacking off a trusted user. Sometimes these attacks do not initially target important or sensitive information, but it is important to remember that all information is important to someone [12]. A well-trained attacker knows how to leverage irrelevant information to eventually get to the information they really want.

2.2.1 The Social Engineering Attack Cycle

Most social engineers follow the same attack cycle consisting of 3 main steps, with the fourth being somewhat optional but highly recommended [13].

1. Research

During this step the attacker will search information on the target (typically a person or particular group of people). This can entail searches on social media for family, likes, dislikes, and anything else they can leverage. If the target exists within a business structure, the attacker will try to learn as much about the business's security practices and culture in order to avoid detection. An adept attacker may first attempt to compromise lower level staff in order to more convincingly go after higher value targets, like managers and department heads [13]-[14].

2. Contact

In this step the attacker will attempt to socially engage with their target. They use the information from their research to try and gain a rapport with the potential victim. Here, the research step proves key: the attacker needs to be sufficiently well informed as to not arouse suspicion when they are making their requests or inquiries [13]-[14]. For every question the victim could raise the hacker must have a response. Furthermore, every response must seem sufficiently convincing and natural. In this way the contact step is like improv, except the end goal is a little different than in a theater. Rather than to entertain, the goal of this show is to create a rapport and a relationship the attacker can use (either in the moment or some point down the line) to manipulate the victim into giving up sensitive data [13]-[14]. This data can include login information or a compromised workstation, all of which are vital jumping-off points for step three...

3. Attack!

   Now that the hacker has set the stage, the true attack is launched. This can be any number of offenses including stealing data, hobbling a company's infrastructure, or even holding the victim's information for ransom [13]-[14].

4. Closing (optional, but highly recommended)

   Closing happens after the attack has been successfully completed. While not every social engineer does this, it is often vital to keeping the attack unnoticed by the victims and even authorities. The closing step is where loose ends are tied up. Digital footprints are erased, and targets of social engineering are hopefully left none the wiser. This also opens the door to reuse a target if necessary. Some types of attacks rely on the target suspecting nothing until it's too late and the attacker is "in the wind" [13]-[14].

2.2.2 Types of Social Engineering Attacks

2.2.2.1 Scareware

Scareware is a variety of social engineering that uses malicious software in combination with fear and anxiety to attempt to force the user to purchase unnecessary software [15]. While this bears similarity to extortion, in a scareware attack the attackers typically do not have real leverage against the user; however, the attackers hope the user does not feel this way. Scareware typically uses pop-ups and messages that may appear as legitimate warnings notifying the target about the (in)security of their system.

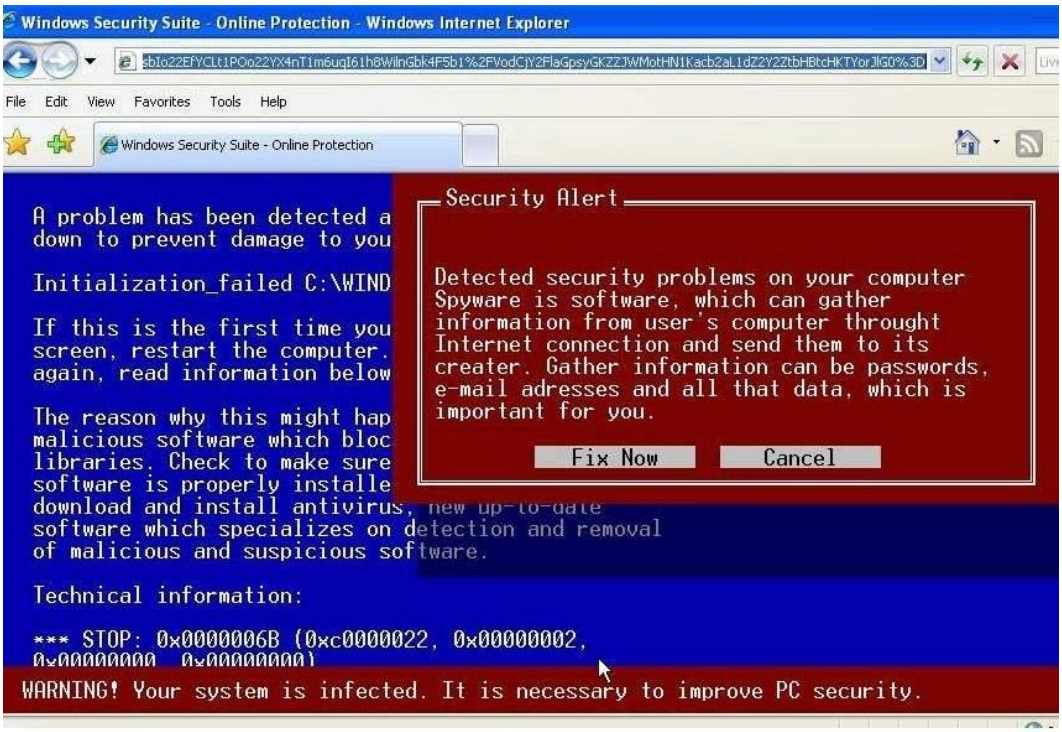


Figure 1
Source: [16]

These programs appear as legitimate by masquerading either as a system (Windows, Linux, Mac) message or as some sort of reputable software like an antivirus program. The programs typically appear on systems that have visited malware infected websites, and as long as they do not contain any spyware, remote access, or ransomware capabilities can be dealt with fairly easily [14]. The goal of the scareware, typically, is to attempt to convince the user their system is infected and/or compromised. The popup then offers the solution to the problem in the form of some paid antivirus or PC repair tool. In reality, the software the user ends up downloading (if they fall prey) will most likely end up being even more aggressive malware or spyware [16].

Scareware is an interesting case of social engineering because it preys on fear and anxiety. It hopes that the user is anxious or fearful enough about their

system being compromised that they will be willing to pay to prevent that without looking into what they're paying for. This can be especially dangerous for inexperienced computer users who are being barraged with messages that make it seem like their computer is on the verge of destruction (in essence, fear-mongering), all while lacking the tools to understand what is really being presented to them.

2.2.2.2 Pretexting

Pretexting is a social engineering attack where the attacker presents themselves as somebody else in order to obtain private information [17]. In these attacks, the impersonation can be done in person, over the phone, or through an unwitting proxy/accomplice. They may present themselves as a construction worker, IT worker, police officer, or any other number of guises that would give them the access they need. This attack relies specifically on the research step– learning exactly what pretext will be the most useful and/or viable in a given situation. The end goal of this attack is to make it seem critical that the attacker retrieve sensitive information and in doing so they may compromise either an individual (think security questions to an account) or a company (think accessing an employee directory) [18].

Typically, the attack ends up going through a single person (i.e. targeting a single employee as a point of entry) regardless of whether the target is an individual or a company. It is always important to remember that for pretexting attacks especially, research is essential. Nowadays, social media allows attackers to do this research more efficiently, and to gain much more information on an

individual, than was possible in the past. In doing so, a well-trained attacker can more easily gain the trust of the person that is chosen as the target. A poorly prepared attacker can quickly compromise their own attack by inadvertently alerting the target that some sort of attack is being perpetrated.

Pretexting attacks prey on trust. They rely both on building trust with their target and then leveraging that trust to acquire information. The target must believe that the attacker, whether over the phone or in person, is trustworthy. This is dependent upon the attacker having a credible background story, or at least being able to convince the target of it. The attacker must also build the trust to last after the fact, at least long enough to allay suspicions to buy time to actually use the sensitive data acquired.

2.2.2.3 Baiting

Baiting consists of leaving devices (typically USBs or SD cards) in a public area to be found and subsequently plugged into a system by an upstanding citizen [14]. The key here is that these devices are typically filled with malware and hacking tools that will run automatically and allow an attacker to steal information on the infected computer. This attack has had tremendous success in the real world. Most famously, the worm known as “Stuxnet” is believed to have infected Iranian nuclear facilities in this way [19]. Stuxnet was a very potent, targeted piece of malware that was distributed through social engineering. Through this avenue, the malware was able to gain access to an air-gapped system that would have been unreachable otherwise. Stuxnet is the perfect example as to how successful–and dangerous–social engineering can be.

2.3 Phishing and its Variations

The following social engineering attacks all derive from the same word, “fishing”. It was originally adapted from f- to ph-, a throwback to “phreaking” (a form of telecommunications hacking beginning in 70s and 80s). And like its aquatic namesake, phishing in all its incarnations is about fishing– this time for information, not fish [20]. A whole host of variations on phishing have arisen over the years, some of them explored below.

2.3.1 Phishing

This is the original variant of this group of attacks and is sometimes used as an umbrella term for all the other attacks discussed in this section. In a common example of a phishing attack, hackers may attempt to impersonate a legitimate company in order to get a target’s login credentials. This is typically done by hosting a fake site that very closely resembles the company they are impersonating. Then an email is sent to the target(s) (thousands of people can be targeted in one attack). From there if a target enters their credential or personal information, they have become a victim of phishing and the attack is successful [20].

2.3.1.1 Mitigations

Typically, a standard phishing attack like the example above relies on the attack material–such as the email and fake website–resembling the real site very closely, with hopes that any inconsistencies go ignored by the targets. When a target overlooks these small inconsistencies, they are much more likely to fall for the scam [21].

These inconsistencies can include URLs that are similar but not identical to the company's official URL. Another red flag is if the navigation buttons of the website/email do not correctly link, or link at all, to where they say they do. One website redirecting to another can also be a telltale sign of a phishing email. Lastly, an email with generic greetings, grammar, and spelling mistakes should also be paid attention to. While not all phishing attacks will have these mistakes, the errors presented here are often telltale signs and should be watched out for.

2.3.2 Spear Phishing

Regular phishing is like casting a large net and hoping somebody falls prey. Spear phishing, on the other hand, involves planning exactly which target to attack and customizing the attack materials to that person [17], [20]. The attack email that is sent out is customized with information about the target: name, employment position, company, work phone, etc. By personalizing the material, the target receives, this attack not only takes advantage of trust but of a perceived social relation that then can exacerbate trust and induce compliance. Again, this attack is made much easier in the current age, when all of the above information is freely available on social media sites (especially LinkedIn for corporate information).

One of the most famous recent examples of this occurred in 2016 when John Podesta, chairman of Hillary Clinton's presidential campaign, was targeted in a spear phishing attack and some very compromising work emails were found and posted online [5]. That event was so life-altering it was felt throughout the United States during the 2016 Presidential Election.

2.3.2.1 Whaling

Whaling is a specific variation of spear phishing that targets high level executives and officers of a company [20]. If a "whale", –think CEO, CFO, vice-president, department heads–is compromised, the attacker gains much greater access than if, say, a customer service representative is compromised. The general style of the attack remains the same as the spear phishing paradigm.

2.3.2.2 Mitigations

The best protection against these types of attacks is keeping employees aware about how these attacks can be carried out. It is also important to encourage employees to avoid publishing information about their personal or professional lives on social media (people end up being very noncompliant on this). Lastly, a good email filtering program can catch a fair number of attempted, less sophisticated attempts.

2.3.3 SMiShing

The core of SMiShing functions the same as regular phishing but instead uses SMS messages (text messages) as the primary attack vector. The attackers again rely on a target to click through to a website, this time presented in an SMS message rather than an email [20].

2.3.3.1 Mitigations

Being wary of unknown phone numbers referencing links or requesting money can thwart this attack fairly effectively. In addition, numbers suspected of this type of activity can and should be reported to the carrier.

2.3.4 Vishing

Vishing is a type of attack where attackers use a phone call placed to the target to try to extract information. These attacks are typically conducted using VoIP servers [20]. They typically require the attacker to be more involved in the information gaining process and also requires they impersonate either a legitimate company or the end user they wish to compromise.

2.3.4.1 Mitigations

In order to protect against vishing attacks, it is important to screen and avoid calls from unknown numbers. However, this is not foolproof as many attackers have devised ways to spoof caller ID in order to appear more legitimate. The attack specifics can vary greatly, so general caution is advised.

2.4 Related Work

2.4.1 Twitter + AI = Great Spear Phishing

John Seymour and Philip Tully, of the company ZeroFOX, used machine learning to create a neural network, "SNAP_R," that when fed Twitter usernames could determine which were the most susceptible to phishing. It did this by using their profiles and post topics. Not only could SNAP_R find a great target, but it could generate content with phishing links and seed them so vulnerable users would see them and potentially click on them as well. Furthermore, SNAP_R does not make one targeted tweet and stop. Rather, it collects the timing of the users' replies in addition to their tweet history to better seed malicious tweets, and can repeat this process to effectively barrage a user [10]. It is important to note that spear phishing campaigns have around a 5 times greater success rate than

non-targeted phishing [10]. And what SNAP_R proved above all is that spear phishing campaigns can be run with little-to-no manual intervention and still be highly effective. The wealth of information available about a potential target online is immense, and many people do not even realize that they have put this information out there. By throwing machine learning into the mix not only can your personal data be used against you, but your seemingly unrelated behaviors can also be used in that same manner (think ad targeting, but much more dangerous).

2.4.2 Phishing Children

Children are a very vulnerable population in today's ever-expanding online world. Children are being introduced to the internet at earlier ages, but unlike their older teenaged counterparts, they are typically not given training on the dangers and scams that can be found online. This is particularly dangerous because children do not necessarily understand what an attacker's goal may be, and may fall prey to an attack and never even notice it. Untrained children were shown to have only about a 60% chance of distinguishing between a legitimate email and a phishing email; however, they did show some improvement after a training intervention [6]. It is also interesting to note that their ability to recognize real emails increased after training and suffered very little decay, while their ability to recognize phishing emails increased after training but subsequently experienced severe decay. This means that when children were presented with a phishing email a few months after training, they performed just as poorly as they

did pre-training, which is consistent with reports from other types of social engineering attack studies [7].

2.4.3 Humans are Bad (at Spotting Lies)

Humans have a penchant for believing they can spot a liar with little to no error. In reality, most people's guess would be just as good if they flipped a coin [8]. This inability to spot liars is not applicable just for those people within the general population, but extends to would-be deception specialists such as police, customs officers, and prison guards as well [9]. For the most part, people are generally awful at spotting lies, though some research like that from Paul Ekman have shown increased deception detection with certain methodologies [9]. However, most of these methodologies require extensive training and practice to be effective in any tangible way. Due to this most people and companies will not receive training on deception detection even though it exists, and are perfect targets for an attacker.

Chapter 3

“Why Do We Take the Bait?!” (It has to do with how we make decisions)

Humans are for the most part intelligent creatures. We have an aptitude for technology unrivaled in the animal kingdom. But nonetheless we are still animals at our core, or at least our brain is. The human brain can be considered a marvel of the modern world; it is a supercomputer we could only dream of building. Animals though we are, our brains are far from simple. When it comes to tasks like facial recognition or complex motor tasks, the human brain can handle these almost without conscious thought, but start talking about probability or complex number theory and we easily spiral into endless confusion. Why is this the case?

The simple answer lies in evolution. Over thousands of years of evolution, the human brain has adapted and evolved for the main goal of all organisms, survival. So, while facial recognition is important to prospering as a social species (i.e. knowing who is a friend and who is an enemy/threat), knowing conditional probability does not really play into survival.

If our social operations are so-called thoughtless, how do humans make decisions?

Though it varies by scenario, social decision-making can typically be broken down into two main processes. Psychologists Keith Stanovich and Richard West put forward the terms System 1 and System 2 to refer to these mental processes, and we will adopt that convention for this paper [22].

3.1 System 1

System 1 is the source of many of the automatic emotional reactions we experience, some of these are expressed in microexpressions [23]. It operates very rapidly, almost instinctively, with little voluntary control or effort [22]. In addition, System 1 "short circuits" System 2 by making a decision before System 2 may even be aware of what is going on, like the decision to remove your hand from a hot stove.

System 1 is really only useful for simple problems that require a negligible amount of attention. Generally, System 1 functions are something everybody is born with: autonomic bodily functions, the ability to perceive the objects in the world, take action to avoid losses of life, limb, or property, and the ability analyze and respond to a social situation swiftly. In special scenarios learned abilities can become System 1 processes, such as driving on an empty road [22].

3.1.1 Heuristics

System 1 has developed a number of heuristics, or mental shortcuts, to short circuit tasks that would be typically relegated to System 2. A lot of these heuristics have to do with what would be considered statistical inference. Examples of these shortcuts include determining the availability of objects in our environment, answering questions with regards to causes of events in the world,

and speculating on the outcome on the possibility of events occurring. The one issue that arises with heuristics is that humans are very bad at statistics, both with System 1 and System 2 [22].

If they can be unreliable, what purpose do heuristics even serve? Well they are not perfect and usually do not lead to the exact right answer, but they give an intuition and a basis for decisions that need to be taken rapidly with minimal lag. And while heuristics are not perfect, they are better than nothing, and can help us survive in situations that require split-second processing.

On the other hand, these same heuristics which make life more manageable also can be the bane of our existence. One heuristic is particularly dangerous– the bias to believe other people are less likely to lie than they are to tell the truth [22]. This specific heuristic can lead to many issues with respect to phishing because when a person receives a dubious letter from the IRS, they have a tendency to believe it is actually from the IRS.

3.2 System 2

System 2 requires that attention be allocated to the mental activities that are being presented. Operations within this system are related to choice, focus, and reasoning [22]. All the processes that are encapsulated in System 2 require attention and are disrupted when attention is broken or reallocated.

While more focused and intentional than System 1, System 2 has some drawbacks stemming from the fact that people have a limited attention budget. For this reason, you cannot drive and do your taxes, or at the very least you should not, but you can drive and have a low maintenance conversation. It has

been proven many times that intense focus is a very real phenomenon and can result inattentional blindness [24].

3.3 1 + 2 = 3, Most of the Time

These two systems in general do not work in isolation. System 1 can change the focus of System 2 and vice versa. For example, System 1 can respond to surprising stimuli in the environment by orienting the eyes in that direction. System 2 is now alerted to pay attention to what is happening in that area [22]. Conversely, System 2 can tell System 1 to scan for features in the environment: a relative at the airport, a name in a list, etc.

System 1 generates suggestions for System 2 of impressions, intuitions, intentions, and feelings. In most cases System 2, with very few or no changes, accepts and implements the suggestions of System 1. They can be in conflict, however, and when that is the case System 2 will allocate attention to the discrepancy.

System 2 will also put attention towards problems System 1 cannot provide an answer for, such as when an event that violates the model of the world of System 1 (like something surprising or shocking) happens, or when the hair-trigger responses of System 1 is about to make a mistake (like telling your thesis advisor he smells). System 2 is key in fighting the impulses of System 1; this is what most people call self-control. Like all well designed systems sometimes the moderation abilities of System 2 can break down or be short-circuited, resulting in System 1 taking control and leading to the person having outbursts of anger or other powerful emotions.

Systems 1 and 2 work in unison and complement each other the majority of the time. They optimize performance and minimize effort, just like systems in a well programmed neural network. But like any network the brain is subject to errors. System 2 acts as the learning function to adjust the biases and automatic responses of System 1 for a situation where the automatic response was not appropriate, in order to streamline and more optimally respond to a potential future encounter.

3.4 How do the systems interact with phishing?

The dialogue between System 1 and System 2 is key to why targets of phishing fall prey to these attacks. It has been mentioned that System 2 is the more rational and logical of the two, whereas System 1 is the more impulsive and rash. Due to this, attackers exploit System 1 and hope System 2 does not react or responds too late. This is possible because System 1 processes thousands of minute decisions at any given time. When it is overwhelmed it can short circuit System 2, especially in cases when it is activated by social, emotional, or biological triggers that are pertinent to immediate survival (survival instincts can also be activated by social situations) [25].

By attacking human emotions and decision-making pathways, attackers want to keep targets' minds working in System 1. Phishers especially abuse principles which can be processed by System 1, some of which have been identified by psychologist Robert Cialdini in what he calls the "Science of Persuasion" [26]. These principles include reciprocation, consistency/commitment, social proof/conformity, likability, authority, and

scarcity. The breadth of principles processed natively by System 1 gives attackers a wide range of topics to exploit in phishing attacks. For example, the threat of an IRS audit (authority), an enticing discount on a hard to acquire product or service (scarcity), or the fact that all users have updated their passwords and you should to (social conformity) are easy claims to make that keep targets thinking using System 1. It is very easy to slip into exclusively System 1 thought processes when scanning emails, especially when a lot of emails that arrive are probably junk. So, phishers try to exploit that particular line of thought. It is always important to think about what you are clicking on in an email and to do the bare minimum to make sure it is real.

3.5 What are these principles?

3.5.1 Authority

Authority is a powerful tool in the social engineer's arsenal. Authority is particularly powerful because the right kind of authority is almost universally respected. However, what kind of authority that right kind of authority is changes on a target-by-target basis. Stanley Milgram's now famous series of shock experiments shows we all have a sense of obedience to authority [26]-[27].

In Milgram's experiment, the authority figure was the boss of the study who urged the teacher to keep delivering painful shocks to the "subject" (an actor not really receiving shocks). Milgram's states, "the extreme willingness of adults to go to almost any lengths on the command of an authority … constitutes the chief finding of [this] study" [27]. His experiment began by looking at how a majority of the German citizenry could be complicit in the Holocaust. While

Milgram's main point was with respect to government, it generalizes to all forms of authority. The fact that most people have an innate sense of obedience to authority is one reason that this principle works so well in phishing attacks.

3.5.1.1 IRS Scams

The IRS is a very popular authority figure for attackers to impersonate to try to gain access to sensitive information. In fact, it is so common the IRS has a long list of resources and ways to identify and report these scams [28]. The IRS is such a perfect example of this authority because the mere mention of their name or an audit can strike fear in almost anybody. As the old adage goes, "they say only two things are certain in life: death and taxes (and only one of those is painless: death)".

3.5.2 Commitment and Consistency

Consistency underlies a lot of System 1 thinking. System 1 thinking relies on your (consistent) general conception of the world. Likewise, the choices you make in that world should be fairly consistent with that conception. Consistency is generally highly valued as a personality trait as well. It offers other people the ability to predict your actions and emotions, which is highly adaptive [26]. In the majority of situations consistency is exactly what is needed, especially to live a normal, well-adjusted life. This default to consistency is exactly what social engineers hope to exploit.

Let's say you are known for donations to cat shelters, and an attacker sends you an email claiming to be from a cat shelter in search of donations. The

possibility you click through that email is very high since it is consistent with your beliefs, ideas, and values.

Consistency lays the groundwork for commitment to take hold. If a person can be convinced to make a commitment, then the principle of consistency will very likely bind them to that commitment [26]. By forcing System 1 to register a new point of consistency in the world through commitment, you can greatly increase the chance that the person in question will remain consistent in the future. Social engineers can really use this to their advantage if they need to reuse a previous target for more information (think– “you agreed to help me earlier, can you help me now?”).

3.5.3 Likability

Unsurprisingly, a person will usually agree to requests made by people that they like [26]. Overall, this is a pretty straightforward and simple principle. There are a number of ways to be likeable, some within an attacker's control and some outside it. First off, attackers can stick to basics: show a smile, be cordial, and do not be too aggressive when trying to get information. Other things that can bolster likability are having similarities in beliefs, background, or lifestyle to the target; being cooperative; and seeming familiar to the target. We gravitate toward things we are familiar with as they provide more consistency in our lives. For this reason, seeming familiar is actually one of the largest ways to come off as likeable. Physical attractiveness can also increase likeability, but most people cannot control this [26].

3.5.4 Reciprocity

Reciprocity is a very useful tool in the social engineer's repertoire because people often feel inclined to help those that help them. This principle states that we should try to repay what another person has given to us [26]. In a fundamental way, this trait helps us strengthen relationships by keeping them from being one-sided. But like any other, it can be manipulated. Picture this: a visher calls and impersonates tech support. After convincing you that they are saving you from the latest zero-day you happily offer up your login credentials so they can solve the problem. Reciprocity comes into play for both large and small gestures and is another principle reinforced by consistency. Reciprocity can push usually unwilling targets to compromise with a phisher because the need to reciprocate can be simply overpowering at times [26].

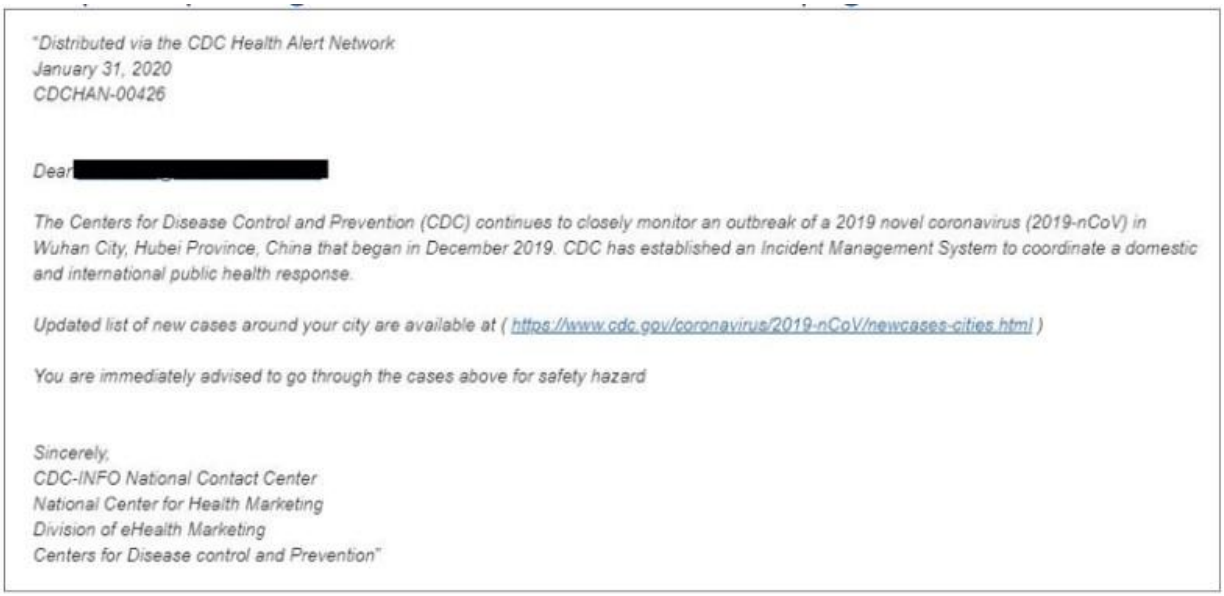
3.5.5 Scarcity and FOMO

Cialdini also introduces the principle of scarcity. At its core, scarcity is the principle that perceived or real scarcity leads humans to place a higher value on the object that is in short supply [26]. The newer term "fear of missing out," or FOMO, is an apt companion to the principle of scarcity. Scarcity of products or information can lead to a dramatic increase in the amount of value that is placed on them. FOMO goes one step further in that it addresses the anxiety people experience when left out of information, social updates, or even products. Coupling scarcity and FOMO together gives you higher valued products that people do not want to miss out on. This increases the scarcity and subsequently the sense of FOMO, causing this cycle to perpetuate itself. Attackers can leverage

both the anxiety and perceived value to disarm targets by offering exclusive access or information. This principle can be leveraged with any real or perceived scarcity and it works even better in moments of crisis. When people are most vulnerable using scarcity and FOMO in conjunction is a powerful tool that can leave even the savviest in an unsavory situation.

3.5.5.1 COVID-19

A perfect example of how attackers use scarcity in times of crisis can be seen during the COVID-19 pandemic. Phishers sent thousands of emails impersonating a wide range of agencies in order to compromise a panicking population that was suddenly shifting to working from home [29]. These hackers abused fear, uncertainty, and even sympathy brought about by the pandemic to try to profit from a very unstable situation [30]. The perceived scarcity of objects like toilet paper and the actual scarcity of testing kits in the early months gave hackers a perfect list of offerings that they could bait potential targets with.



"Distributed via the CDC Health Alert Network
January 31, 2020
CDCHAN-00426

Dear [redacted]

The Centers for Disease Control and Prevention (CDC) continues to closely monitor an outbreak of a 2019 novel coronavirus (2019-nCoV) in Wuhan City, Hubei Province, China that began in December 2019. CDC has established an Incident Management System to coordinate a domestic and international public health response.

Updated list of new cases around your city are available at (https://www.cdc.gov/coronavirus/2019-nCoV/newcases-cities.html)

You are immediately advised to go through the cases above for safety hazard

Sincerely,
CDC-INFO National Contact Center
National Center for Health Marketing
Division of eHealth Marketing
Centers for Disease control and Prevention"

Figure 2
Source: [29]

3.5.6 Social Proof and Social Conformity

Social proof and social conformity stand very closely together. Humans make decisions on what to do by determining what is correct. Often, they determine what is correct by finding out what other people think is correct, a principle known as social proof [26]. Social proof gives people the confidence necessary to make many decisions while operating under the mindset of, “it works for them, it should work for me too”. People also look for validation in their peers. Often, they feel indirectly ostracized by not conforming to what others are doing, a phenomenon known as social conformity [26]. This can be even more pronounced when the person that is being looked towards as an example has some sort of influence. This is why Instagram influencers and celebrities that go out and buy a product or start a hashtag challenge can begin a trend that catches on very quickly. Humans are social creatures that crave acceptance and conforming to the surrounding society is a way of finding that social validation [25].

3.6 Does phishing work?

Does phishing work? After all this, the simple answer is clearly yes. A more nuanced answer, however, is that phishing does work generally, but its degree of effectiveness is determined by the specific type of phishing variant, the technique used, and the demographic targeted. Older adults were affected more by reciprocity while younger adults responded more frequently to scarcity [31]. In a not-so-shocking turn of events, authority was a very powerful motivator across the board regardless of age. As mentioned earlier, children were also very susceptible to clicking on phishing emails [6].

3.7 Why does phishing work?

Phishing works thanks to how the human mind is wired, and because it preys mainly on human emotion. The fact that humans are social animals ready to please, and the fact that anxiety can take over our lives, are both utilized for nefarious means by hackers. The way our brain is designed to handle emotion and social interaction gives hackers a large attack vector and very little mitigation ability. Underlying Cialdini's "Principles of Persuasion" is emotion. The key emotions that social engineers exploit are three of the six basic ones: fear, anger, and disgust [32]. These emotions are the negative emotions and are key because they can induce action more swiftly and instinctually than positive emotions [33].

Found in at least half of the principles listed above (authority, scarcity/FOMO, and social conformity), fear is by far the most easily manipulable emotion. Anger and disgust have smaller impacts but still do come into play; however, fear definitely takes the lead. Conversely our attraction to commitment, consistency, and reciprocity are less emotionally charged; because of this they take advantage of another underlying factor we will discuss later. Fear is one of the few emotions that can trigger the human survival instinct every time. Positive emotions on the other hand, and even lesser negative emotions like anger and disgust, do not trigger that response nearly as reliably (or at all). Fear induces "what if" thinking, triggers chemical cascades that increase stress and anxiety (fight or flight response), and decreases critical thinking. When afraid, humans can be susceptible to phishing attacks that they might otherwise not have fallen for if they were thinking clearly [22]. In most scenarios the human survival

instinct would be life-saving, but in this problem faced in our modern world it is much more detrimental than anything else.

Chapter 4

Trusting Trust and Authority

Fear is a large part of the story, but there is another player involved: trust. An excellent place to study the role of trust in phishing is with spear phishing.

Spear phishing is regarded as more dangerous and proven more effective than your standard phishing technique [10]. Why is this?

The main difference between standard phishing and spear phishing is the use of a tailored message with either personal information or something that speaks to the target's values, likes, or personality. The fact that the message is tailored to the target makes it seem all the more trustworthy. Since it seems more trustworthy, the target is more likely to click through. Once they do that, they are no longer a target, they are a victim.

Trust is the basis for our everyday interactions and is essential in almost everything we do. For that reason, it is an incredibly powerful manipulation tool. As Rotter says, trust is:

> The entire fabric of our day-to-day living, of our social world, rests on trust – buying gasoline, paying taxes, going to the dentist, flying to a convention – almost all our decisions involve trusting someone else. [34, p. 443].

The fact that trust is so crucial to the very fabric of society, gaining trust is the social engineer's equivalent of gaining a skeleton key: any door they want can now be unlocked. Since trust is so crucial to everyday life, most people default to trusting unless given a reason not to [35]. The issue with this is that by the time they have a reason not to trust someone, most social engineers will have already gained what they wanted

Due to this default to trust and the fact that humans are very bad at deception detection, most people naturally assume that people tell the truth [8]. And for the most part this assumption works pretty well. However, the same pattern emerges here as it has with previous principles. When something is adaptive and useful most of the time, the times when it is neither of those things are the perfect chances for hackers to take advantage. Since humans assume truth-telling as the default, when they are faced with a call from tech support, they assume that it is tech support and not someone impersonating tech support who wants to steal their identity. Trust is such a powerful motivator and has a significant hand in the principles (likability, commitment, consistency, and reciprocity). These principles are less emotionally charged and as such need another motivator to rationalize why they work. Trust is that motivator. When you like someone, you trust them more. If someone has given you assistance you trust them more and reciprocate. You usually only make commitments to those you trust (exceptions do exist). When you trust someone, you are more inclined to comply with them, paving the way for the data breach. Now we have a

rationalization for most of the principles of persuasion, but that still does not complete the story.

Trust also plays a role in why some principles are more effective than others. To see how, we revisit the study by Oliviera et al., where researchers saw high levels of response across all age ranges for phishing emails that used authority [31]. Why is authority so special? Authority combines, more so than any other of the principles, fear and trust in perfect harmony. On some level everybody fears the police and the IRS. At the same time, they trust them because they have been told that these organizations are trustworthy. It is this perfect balance between fear and trust that makes authority one of the most powerful principles to persuade people to comply. Does anybody really want to be on the wrong side of the IRS?

The positioning of trust also explains why the other principles are not as effective. This comes about because the remaining principles can all feature emotion and trust, but they do not balance them in the right proportions. Let us consider likability. It can involve positive emotions and trust, but if fear is added to the mix the hacker has just defeated the purpose of being liked in the first place. We can also look at scarcity/FOMO, we see that fear can really motivate. But is there trust utilized here? When people are overflowing with anxiety, their trust tolerance goes down. Following that, trust really will not be the focal point when scarcity/FOMO is being used to persuade a target. The proportions of trust and fear in this principle are off or just plain incompatible. Trust is a very powerful tool that when used correctly can be particularly dangerous.

Chapter 5

Why Should You Care?

Maybe you see these examples of phishing and say, “yeah, but that only happens to idiots. It will not happen to me!” This is not a measure of intelligence, however. Phishing is a systemic issue that anyone can fall prey to, be they a well-trained IT professional or a naïve college student. A lapse in judgement, a misclick, or gnawing curiosity can make a fool out of anybody.

Phishing, like other social engineering attacks, abuses flaws in the overall design of the brain. As much as everybody would like to believe that they are smart enough to avoid falling for these types of attacks, countless intelligent, well informed individuals have been victims to phishing and related attacks. No one is immune to this issue. In addition, when a malicious actor is intent on compromising a particular person or company, they can continually barrage the target until one small mistake is made [36]. If they do not trick you the first time, they can try again and again until you slip up. Even the most minute mistake from an intelligent person can render an attack successful.

It is hard to be hyper aware–and secure–in every aspect of our online activity, especially as everything in our lives is becoming more reliant on

technology. Nearly everyone has some company they give their data to, whether by choice or obligation. It has become a necessity in this day and age. For that reason, the end user is not the only one that has to be hyper aware– the companies we give our data to must also be compliant. Every individual worker they employ has to be proactive and practice good security religiously. But once you include that many points of failure into a system, one of them is almost guaranteed to fail. It might not be today or tomorrow, but it will happen. And more often than not, it is the end users that pay the larger price.

5.1 Who is affected?

5.1.1 43% of all people

In Oliveira's study, 43 percent of participants clicked the link in the "phishing" email and at least once, and 11.9 percent clicked more than once [31]. This striking number implies that only slightly less than half of all people are susceptible to phishing attacks. Of course, this susceptibility is not evenly spread. Factors like age, gender, education, and cognition all change the way people respond to phishing [31]. Older populations, specifically women older than 62, were significantly more susceptible to phishing emails than any other populations. The study links this to the cognitive decline that accompanies old age. As cognition declines, so does the ability to recognize deception [37]. But while there are certain groups of people found to be more susceptible, people of many demographics fell victim to the scam (and some even fell victim multiple times). No group was found to be immune.

5.1.2 Celebrities

For an example that anyone can be phished we may turn to Barbara Corcoran. She is by all means a very successful businesswoman, and serves as a judge on ABC's "Shark Tank". But regardless of her intelligence and success, a fake invoice for real estate renovation costs allowed hackers to scam her out of $388,700 [38]. Perhaps she fell into processing with System 1– she automatically registered the invoice as real because it was a routine expense for her, given her involvement in real estate. By the time her bookkeeper noticed the discrepancy the money and the attacker were both gone. If a successful woman worth over $80 million dollars can be phished, what hope for immunity do the rest of us have?

5.1.3 Corporations

While individuals are often the targets of phishing attacks, corporations also find themselves as the target of these attacks as well. In these cases, not only is the corporation a target, but the data of hundreds of thousands of people could also be at risk.

5.1.3.1 Sony

In 2014 Sony Pictures was attacked by malicious actors that subsequently released business agreements, financial documents, and Sony employees' information [39]. The act was blamed on North Korea, but whoever the attacker was is irrelevant. Access is believed to have been gained using spear phishing emails impersonating Apple targeted to Sony employees. It only took one employee to click the link to compromise 100 terabytes of data [39]. While Sony was massively affected, the thousands of employees whose personal information

(names, DOB, SSN, etc) was leaked online for anybody to use were the ones truly left out in the cold. This is a perfect example as to why even if you as an individual practice immaculate security, you are still not immune from phishing.

5.1.3.2 Yahoo

In another famous case from 2014, Yahoo fell victim to a massive data breach that was not limited to its employees. This data breach endangered around 500 million users by exposing the usernames, passwords, phone numbers, emails, and cryptographic values (hashes) associated with their accounts [40]. The attack relied on a spear phishing campaign that targeted semi-privileged employees. It is reported that one employee fell for the email granting the attacker access to the database that was later leaked. It only took one person to succumb to put 500 million at risk. With a ratio like that phishing should worry all of us.

5.1.4 The United States Government

This paper references the John Podesta (Hillary campaign) email leak earlier, which is government-related but nonetheless involved a corporation. But do governments have an easier time avoiding phishing? This, sadly, is wishful thinking. An example can be found in the US territory of Puerto Rico, whose government lost $2.6 million due to a phishing attack in January of 2020 [41]. According to reports, the government agency received an email that claimed a change had occurred to a bank account tied to remittance payments. The agency complied with the email and ended up making the transfer requested to the fraudulent account. It almost seems too simple, but the attackers made off with $2.6 million dollars because of a single email no one fact checked.

All of the examples detailed here serve to prove that no one is safe. A dedicated hacker could craft a phishing attack that would make even the most security-oriented person thinks about clicking on it. Not all scams look like the Nigerian 419 scam, but they are all equally as dangerous. In addition, a well-engineered phishing scheme can affect an individual directly or indirectly, so in that sense it does not matter how security-oriented to the individual themselves is. The fact that no one can avoid is why we should all care about phishing. It does not matter if you are an individual, a company, or even a government: the social engineer is fighting against a single human in every case.

Chapter 6

Why Phishing is Around to Stay

There is no magical panacea to solve and protect us as a society from the dangers of phishing. There is no program that can defend against every phishing attack and there is no quick solution to the issue of being human, either. So, no matter how you feel about it, as a technologically advanced society, phishing is here to stay for the foreseeable future.

6.1 Deception Is Around to Stay

We cannot eliminate deception. Humans have been deceiving each other offline for thousands of years, so continuing online was an inevitable next step. Phishing is one of the many forms that deception has taken in cyberspace, and as long as people remain awful at detecting deception, this practice will stick around.

6.2 Humans are Forgetful

You can train humans to identify phishing attempts and they will become more conscious of the threats. The issue, however, is that they will eventually forget what they learned [7]. Over time they will slip into System 1 defaults and the effects the training had on their phishing identification skills will be

negligible. Phishing is not one of things we are designed to worry about and because of that it will be hard to get rid of or protect against.

6.3 Humans are Hardwired for Failure

The shockingly real truth is that as humans, evolution has put us at a few thousand years' disadvantage when it comes to phishing. While we are psychologically well adapted for most survival situations, it is not our fault that we fall prey to these attacks that we were never designed to deal with. It is not our fault that we are human. With that being said, it does not matter why our brains are wired the way they are or what cosmic power caused it to be that way. Our brains are not designed to handle phishing well in their current state and we have yet to figure out a way to alter that physically in any significant and adaptive manner. Given this, we have to live with our brains as they are, with all the benefits and disadvantages that come with them.

6.4 Technology Evolves for Everyone – Including the Bad Guys

There is currently no technology that can perfectly filter phishing attacks. And while machine learning has helped tremendously for email phishing detection, it has not progressed in a meaningful way to prevent other phishing variations like vishing and SMiShing. On top of that, machine learning is only somewhat decent at filtering very blatant phishing attempts. Phishing is an ever-changing threat and we have yet to find a technological model that can predict and proactively act against these shifts.

Even in the cases where we have found a defense for phishing, hackers have adapted their schemes to it with little issue. Targeted spear phishing campaigns clearly illustrate this principle. A targeted spear phishing email is designed to mimic real emails as closely as possible to compromise a very specific target. This is a perfect example of an adaptation that has been created to circumvent traditional email filtering algorithms meant to discourage phishing.

Advances in technology are not exclusive to anybody. For every defensive technological advance there is an advance in the methods and techniques used to hack past it. People have made careers, both legitimate and illegal, out of circumventing and defeating the newest cyberdefenses. A new defense against phishing will likely be met with a creative and motivated number of people looking for a way to circumvent it. We can create new technology to defend us online, but we cannot prevent attackers from adapting or creating methods to get around that same defense.

The above may seem to paint a very bleak picture for the human race. We are a group of unadaptable, foolish, and gullible individuals that are just destined to fall for deception. While that may read like a disparaging account of human ability, it is meant to discuss why phishing in all its incarnations is around to stay in our technological lives. By discussing why, we can now provide solutions that can take into account human limitations and tendencies. In doing so we can create more useful and effective mitigations against phishing, as well as against social engineering more broadly.

Chapter 7

Mitigations for Phishing

> The lack of a security mindset explains a lot of bad security out there … Teaching designers a security mindset will go a long way toward making future technological systems more secure. That part's obvious, but I think the security mindset is beneficial in many more ways. If people can learn how to think outside their narrow focus and see a bigger picture, whether in technology or politics or their everyday lives, they'll be more sophisticated consumers, more skeptical citizens, less gullible people [42].

Before we begin, it is important to remember that mitigation is not one-size-fits-all. What works to remind one person to act in a secure manner might not for the next person. In addition, we see that people have a tendency to forget previous training over time rendering them just as vulnerable and possibly even more so because it may give them a false sense of security. Nonetheless, there are definitely steps that the collective we (researchers, developers, companies, and the average person) can do to limit the amount of damage possible. While the proposed mitigations may not be 100% effective, they are useful and much better than no plan whatsoever.

7.1 Behavioral and Societal Mitigations

An important first acknowledgement to make before discussing any type of behavioral change is that potential interventions should not attempt to change the human brain. These attempts will not work and are not worth the time. Instead it is important to implement these behavioral mitigations through learning with the brain that we are given.

7.1.1 A Security Mindset

From a behavioral standpoint, one of the most important things we can do is foster a "security mindset" [42]. What that means is we should practice thinking like an attacker and imagine how objects, programs, and humans can be made to fail. But a security mindset is something that needs to be developed over time. As a society we should not discourage seeking flaws out, nor should we enforce such a rigid status quo. A security mindset is antithetical to complacency with the norm. Developing a security mindset is a goal that must be continually worked towards, there is no easy way out. While certain instincts are inflexible, our point of view is surprisingly malleable given enough time and practice.

7.1.2 Trust but Verify

As a society, we should also encourage adults and children to have a healthy amount of skepticism and distrust. This is not to imply we should devolve into a paranoid society that does not trust one another and always assumes the worst, as that would be wildly counterproductive. We want a healthy amount of skepticism in our lives, not a dysfunctional amount. An important extension of this is our willingness to concede to authority due simply to fear and trust. By

instilling as a society a sense of respect for authority, but also teaching that we have the right to be skeptical and ask for proof, we can massively decrease the power authority has on our decision making. This is extremely important as authority was one of the most powerful motivators for all age demographics to fall for phishing attacks [31]. “Trust but verify” would be a good phrase to encapsulate the behavior we should practice and what we as a society should work towards.

7.1.3 Common Sense

Sometimes in our current world common sense does not seem all that common. Maybe the basics have been forgotten, or maybe they were never even taught. Staying cognizant of a few reminders would help us (even those of us for whom common sense is more of a struggle) stay alert to potentially dangerous situations:

- If an offer sounds too good to be true, it probably is. It may sound tempting in the moment, but it's very unlikely you just won a new iPad.
- Do not automatically trust strangers, both online and in person. You would be wary of a package or a letter from someone you did not know and it should be the same for messages you receive
- If you are suspicious, confused, or worried about a message you received you have resources right at your fingertips. Use them! Google the details, ask around online, verify contact info, and do your research

- Report suspicious behavior and messages. You do it in your neighborhood and you should do it online too

7.2 Technological Mitigations

Technological mitigations build off of the key points of the behavioral and societal mitigations. We want to implement these technological mitigations with a security mindset, the principle of trust but verify, and some common sense. While as is mentioned above technology is constantly being circumvented, it is never a bad thing to make things harder for hackers. The more difficult it is for malicious attacks to come your way, the fewer that will come to you.

7.2.1 Enable 2FA

If a company or service offers two factor authentication, take advantage of it. It is definitely not perfect, but is a good second line of defense if someone was able to phish access to an account. It is also not necessarily convenient, but that is the point. It is meant to verify that you, the owner of the account, is actually trying to access the account. So, while it may take a few more seconds to log in, you will be grateful for 2FA after you "accidently" clicked that suspicious link.

7.2.2 Update Your Systems

Do not put off system and security updates! This means updating your software when the new update comes out. Keeping your system and security updated is very important because it lessens the number of attack vectors present, but also helps to protect you if you did end up downloading an unsavory program on accident. If for some bizarre reason you decide to only ever update one thing,

at least make sure your malware/antivirus software is constantly updated and set to scan your computer automatically.

7.2.3 Lock Your Devices

Make sure you put passwords on all your electronic devices. You should not leave them unattended and unlocked where someone could gain physical access to them, either. By doing this you decrease the risk of an attacker using physical access to bypass additional security measures like 2FA.

7.2.4 Keep Working on Defense

It is important to continually work on defensive programs and algorithms to try to fight against phishing. Phishing is a very difficult problem with a lot of variation, which makes it even harder to solve. But novel methods must be continually developed to fight against attacks like phishing. The attackers will not stop innovating, so the defenders cannot stop either.

New technologies should not be limited to identifying and filtering of phishing emails. Exploring ways in which to verify the sender to messages is key to fighting spoofing. We also should expand into more novel and easier to implement methods of defending against SMiShing and vishing. We cannot exclusively focus on email phishing because these are avenues of attack that are just as dangerous.

7.3 Corporate Mitigations

Corporations have the responsibility to both their employees and their users to implement and respect good cybersecurity practices. They should focus

on the security habits of the people who work for them in order to implement effective methods of defense. It always comes back to the people!

7.3.1 Enable (Corporate) 2FA

Since corporations are responsible for security, it is their responsibility to force 2FA on their users and employees. It is a measure that is very simple but can prevent a large number of threats where attackers try to impersonate the owner of a corporate account. It is not foolproof, but it is a great buffer for a lot of social engineering attacks.

7.3.1 Train (and Re-Train) Your Employees

Training should not attempt to change the way people act at a core level. Instead, it should leverage the tools that typically work against them to work for the company. As we have seen, training to increase awareness and recognition of phishing does work, but we eventually devolve to pre-training System 1 behavior over time. The key to this is continually maintaining and re-training your employees. System 1 can be taught new defaults; it just takes a while. By continually reinforcing the training, System 1 can begin to learn how to parse and identify phishing subconsciously without falling prey. It is similar to learning to drive: when most people start driving, System 2 has to be in control 100% of the time. As experience is gained, System 2 can teach and relegate tasks to System 1, and because of this most people are able to operate in System 1 to a large extent while driving. This means System 1 has learned to perform a new task and knows when to refer to System 2, which is exactly what we want for phishing detection. But the key is training until System 2 has changed the way System 1 operates.

System 2 needs sufficient time to make the necessary adjustments to the beliefs and understandings that System 1 has of the world.

7.3.3 Provide an Avenue for Reporting

If we want people to report suspicious behavior, we need to give them a way to do so. Within a company it is important to provide your employees and product consumers with an easy and hassle-free way to report phishing and other social engineering attacks. For this to be effective, of course, employees must also be willing to collaborate and report potential attacks. They must receive training on how to do so and the process needs to be simple and straightforward. If a person must write a 20-page dissertation on what happened in order to report, no one is going to report anything. What a streamlined avenue for reporting provides for the information security (InfoSec) team is the ability to be proactive and check for breaches or potential compromises. It also gives the InfoSec team the ability to check for patterns, track who is being targeted, and monitor potential threats over time.

7.3.4 Be Prepared Not Negligent

A company has the responsibility to implement good security practices. Because of this they should be prepared for the worst case. In the event of a compromise, the company should have a well-prepared plan to mitigate the damage (like the ability to reset all passwords immediately). They should practice good data storage and encryption. They should not keep sensitive user or employee information on the same server as other business information. There

should be some compartmentalization to limit exposure in the case of a successful breach.

In addition, these companies should require their employees and users to keep good security practices (2FA, strong passwords, changing passwords, etc). And while this may be mildly inconvenient to users the benefits are definitely worth it.

Chapter 8

Conclusion

Our brains have hardwired us for failure when it comes to phishing. As defeatist as that might sound, it is nonetheless the truth. Our psychology has put us in the position where we are very susceptible to these security breaches. Some of us are affected more by different styles of persuasion, but emotion and trust are the main players in any phishing attack.

The fact that so many of us can fall prey to phishing is nothing to be ashamed about. It is the price, so to speak, of being human. While we might not want to live with either, both phishing and deception more generally are around to stay. However, we as humans are not helpless and we can fight back. Our brains may be hardwired to fall for phishing, but they are also malleable. We can be taught new biases and behaviors. It just takes some time and effort. The collective we (researchers, developers, companies, the average person) must all play a role in changing our society and fighting against phishing and related attacks. Everybody must work towards a society that has a security mindset, can trust but verify, and uses a little bit of common sense. While phishing might be around to

stay, if we can pull these steps off we can significantly decrease the amount of damage it causes.

References

[1] KnowBe4, "History of Phishing," *Phishing*. [Online]. Available: https://www.phishing.org/history-of-phishing.

[2] A. Elledge, "Phishing: An Analysis of a Growing Problem," *SANS Institute InfoSec Reading Room*, Jan-2007.

[3] "Advance Fee Schemes," *FBI*, 15-Jun-2016. [Online]. Available: https://www.fbi.gov/scams-and-safety/common-scams-and-crimes/advance-fee-schemes.

[4] "2019 Phishing and Email Fraud Statistics," *retruster*. [Online]. Available: https://retruster.com/blog/2019-phishing-and-email-fraud-statistics.html.

[5] "Threat Group-4127 Targets Google Accounts," *Secureworks*, 26-Jun-2016. [Online]. Available: https://www.secureworks.com/research/threat-group-4127-targets-google-accounts.

[6] E. Lastdrager, I. C. Gallardo, P. H. Hartel, and M. Junger, "How effective is anti-phishing training for children?," *Proceedings of the Thirteenth USENIX Conference on Usable Privacy and Security*, 01-Jul-2017. [Online]. Available: https://dl.acm.org/doi/10.5555/3235924.3235943.

[7] J.-W. Bullee, L. Montoya, M. Junger, and P. H. Hartel, "Telephone-based social engineering attacks: An experiment testing the success and time decay of an intervention," *Proceedings of the inaugural Singapore Cyber Security R&D Conference (SG-CRC 2016)*, Jan-2016. [Online]. Available: https://doi.org/10.3233/978-1-61499-617-0-107.

[8] M. Hartwig and C. F. Bond, “Why do lie-catchers fail? A lens model meta-analysis of human lie judgments.,” *Psychological Bulletin*, vol. 137, no. 4, pp. 643–659, Jul. 2011.

[9] A. Vrij and G. R. Semin, “Lie experts beliefs about nonverbal indicators of deception,” *Journal of Nonverbal Behavior*, vol. 20, no. 1, pp. 65–80, Mar. 1996.

[10] J. Seymour and P. Tully, “Weaponizing data science for social engineering: Automated E2E spear phishing on Twitter,” *Blackhat*, 04-Aug-2016. [Online]. Available: https://www.blackhat.com/docs/us-16/materials/us-16-Seymour-Tully-Weaponizing-Data-Science-For-Social-Engineering-Automated-E2E-Spear-Phishing-On-Twitter-wp.pdf.

[11] R. Moir, “Defining Malware: FAQ,” *Microsoft Docs*, 01-Apr-2009. [Online]. Available: https://docs.microsoft.com/en-us/previous-versions/tn-archive/dd632948(v=technet.10)?redirectedfrom=MSDN.

[12] J. Fox, “The Dark Arts Of Social Engineering,” *SANS Security Awareness Summit 2018*, 23-Oct-2018. [Video] Available: https://www.youtube.com/watch?v=FvhkKwHjUVg.

[13] “The Attack Cycle,” *Security Through Education*. [Online]. Available: https://www.social-engineer.org/framework/attack-vectors/attack-cycle/.

[14] B. Dobran, “Understanding The Latest Tactics & Threats In Social Engineering For 2020,” *PhoenixNAP Global IT Services*, 19-Dec-2019. [Online]. Available: https://phoenixnap.com/blog/what-is-social-engineering-types-of-threats.

[15] "What is Scareware?," *Kaspersky*. [Online]. Available: https://usa.kaspersky.com/resource-center/definitions/scareware.

[16] "What is Scareware?," *FraudWatch International*, 08-Mar-2017. [Online]. Available: https://fraudwatchinternational.com/expert-explanations/what-is-scareware/.

[17] "Social Engineering," *Imperva*. [Online]. Available: https://www.imperva.com/learn/application-security/social-engineering-attack/.

[18] "Social Engineering on Social Media," *Cyber Security Agency*, 12-Jun-2017. [Online]. Available: https://www.csa.gov.sg/gosafeonline/go-safe-for-me/homeinternetusers/social-engineering-on-social-media.

[19] D. Kushner, "The Real Story of Stuxnet," *IEEE Spectrum: Technology, Engineering, and Science News*, 26-Feb-2013. [Online]. Available: https://spectrum.ieee.org/telecom/security/the-real-story-of-stuxnet.

[20] D. Bisson, "6 Common Phishing Attacks and How to Protect Against Them," *tripwire*, 07-Oct-2019. [Online]. Available: https://www.tripwire.com/state-of-security/security-awareness/6-common-phishing-attacks-and-how-to-protect-against-them/.

[21] L. Kane, "There's a reason Nigerian scammers are so obvious in their emails," *Business Insider*, 28-May-2014. [Online]. Available: https://www.businessinsider.com/why-nigerian-scam-emails-are-obvious-2014-5.

[22] D. Kahneman, *Thinking, fast and slow*. New York: Farrar, Straus and Giroux, 2015.

[23] P. Ekman, "Lie Catching and Microexpressions," in *The Philosophy of Deception*, C. Martin, Ed. Oxford: Oxford Univ. Press, 2009, pp. 118–138.

[24] C. F. Chabris and D. J. Simons, *The invisible gorilla: and other ways our intuitions deceive us*. New York: MJF Books, 2012.

[25] A. M. Evans and J. I. Krueger, "The Psychology (and Economics) of Trust," *Social and Personality Psychology Compass*, vol. 3, no. 6, pp. 1003–1017, 2009.

[26] R. B. Cialdini, *Influence: the psychology of persuasion.* New York: Collins, 2007.

[27] S. Milgram, *Obedience to Authority: An Experimental View*. New York, NY: Harper Perennial Modern Thought, 2019.

[28] "Suspicious e-mails and Identity Theft," *Internal Revenue Service*. [Online]. Available: https://www.irs.gov/newsroom/suspicious-e-mails-and-identity-theft.

[29] S. Symanovich, "Coronavirus phishing emails: How to protect against COVID-19 scams," *Norton*. [Online]. Available: https://us.norton.com/internetsecurity-online-scams-coronavirus-phishing-scams.html.

[30] P. Crosman, "Coronavirus phishing scams proliferate," *American Banker*, 30-Mar-2020. [Online]. Available:

https://www.americanbanker.com/news/coronavirus-phishing-scams-proliferate.

[31] D. Oliveira, H. Rocha, H. Yang, D. Ellis, S. Dommaraju, M. Muradoglu, D. Weir, A. Soliman, T. Lin, and N. Ebner, "Dissecting Spear Phishing Emails for Older vs Young Adults," *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Feb. 2017.

[32] T. Dalgleish, M. Power, and P. Ekman, "Basic Emotions," in *Handbook of Cognition and Emotion*, Malden, MA: Wiley Interscience, 2005, pp. 45–60.

[33] C. Eben, J. Billieux, and F. Verbruggen, "Clarifying the Role of Negative Emotions in the Origin and Control of Impulsive Actions," *Psychologica Belgica*, vol. 60, no. 1, pp. 1–17, Jan. 2020.

[34] J. B. Rotter, "Generalized expectancies for interpersonal trust.," *American Psychologist*, vol. 26, no. 5, pp. 443–452, 1971.

[35] D. Dunning, J. E. Anderson, T. Schlösser, D. Ehlebracht, and D. Fetchenhauer, "Trust at zero acquaintance: More a matter of respect than expectation of reward.," *Journal of Personality and Social Psychology*, vol. 107, no. 1, pp. 122–141, 2014.

[36] "Social Engineering | Cyber Security Crash Course," *F-Secure*, 07-Mar-2018. [Video] Available: https://www.youtube.com/watch?v=hZbgnFeXlr0.

[37] N. C. Ebner, D. M. Ellis, T. Lin, H. A. Rocha, H. Yang, S. Dommaraju, A. Soliman, D. L. Woodard, G. R. Turner, R. N. Spreng, and D. S. Oliveira, "Uncovering Susceptibility Risk to Online Deception in Aging," *The Journals of Gerontology: Series B*, vol. 75, no. 3, pp. 522–533, 2018.

[38] C. Brito, "'Shark Tank' star Barbara Corcoran loses $388,700 in phishing scam," *CBS News*, 27-Feb-2020. [Online]. Available: https://www.cbsnews.com/news/barbara-corcoran-loses-388700-dollars-phishing-scam-shark-tank/.

[39] K. Zetter, "Sony Got Hacked Hard: What We Know and Don't Know So Far," *Wired*, 03-Jun-2017. [Online]. Available: https://www.wired.com/2014/12/sony-hack-what-we-know/.

[40] D. Volz, "Yahoo says hackers stole data from 500 million accounts in 2014," *Reuters*, 23-Sep-2016. [Online]. Available: https://www.reuters.com/article/us-yahoo-cyber/yahoo-says-hackers-stole-data-from-500-million-accounts-in-2014-idUSKCN11S16P.

[41] C. Fisher, "Puerto Rico's government lost $2.6 million to a phishing scam," *Engadget*, 13-Feb-2020. [Online]. Available: https://www.engadget.com/2020-02-13-puerto-rico-government-loses-millions-phishing.html.

[42] B. Schneier, "Inside the Twisted Mind of the Security Professional," *Wired*, 04-Jun-2017. [Online]. Available: https://www.wired.com/2008/03/securitymatters-0320/.