How social engineers use persuasion principles during vishing attacks

Keith S. Jones, Miriam E. Armstrong and McKenna K. Tornblad
Department of Psychological Sciences, Texas Tech University, Lubbock, Texas, USA, and
Akbar Siami Namin
Department of Computer Science, Texas Tech University, Lubbock, Texas, USA

Received 14 July 2020
Revised 2 October 2020
Accepted 12 October 2020

Abstract

Purpose – This study aims to examine how social engineers use persuasion principles during vishing attacks.

Design/methodology/approach – In total, 86 examples of real-world vishing attacks were found in articles and videos. Each example was coded to determine which persuasion principles were present in that attack and how they were implemented, i.e. what specific elements of the attack contributed to the presence of each persuasion principle.

Findings – Authority (A), social proof (S) and distraction (D) were the most widely used persuasion principles in vishing attacks, followed by liking, similarity and deception (L). These four persuasion principles occurred in a majority of vishing attacks, while commitment, reciprocation and consistency (C) did not. Further, certain sets of persuasion principles (i.e. authority, distraction, liking, similarity, and deception and social proof; , authority, commitment, reciprocation, and consistency, distraction, liking, similarity and deception, and social proof; and authority, distraction and social proof) were used more than others. It was noteworthy that despite their similarities, those sets of persuasion principles were implemented in different ways, and certain specific ways of implementing certain persuasion principles (e.g. vishers claiming to have authority over the victim) were quite rare.

Originality/value – To the best of authors' knowledge, this study is the first to investigate how social engineers use persuasion principles during vishing attacks. As such, it provides important insight into how social engineers implement vishing attacks and lays a critical foundation for future research investigating the psychological aspects of vishing attacks. The present results have important implications for vishing countermeasures and education.

Keywords Persuasion, Social engineering, Attacks, Persuasion principles, Social engineering attacks, Vishing

Paper type Research paper

Introduction

Social engineering (SE) is convincing someone to act in a way that is not in their best interest (Hadnagy, 2011). SE *attacks* target human victims, aiming to convince them to give the attacker access to restricted information systems or divulge secure information (Gupta and Agrawal, 2012).

Common SE attacks include phishing and vishing (The Social Engineering Framework (SEF), 2019). Phishing occurs when an attacker sends electronic messages to a group of

This research was supported by the National Science Foundation (NSF) under award # 1723765. Opinions, findings and conclusions are those of the authors and do not necessarily reflect the views of NSF.

2056-4961
DOI 10.1108/ICS-07-2020-0113

people in an attempt to fool victims into divulging personal information (Maggi, 2010). For example, an email appearing to be from a university might ask the recipient to click a link to update their password. Vishing (voice phishing) occurs when an attacker attempts to obtain information from a victim over the phone (Maggi, 2010). These attacks can be launched using phone numbers and personal information mined from caller ID and social media applications that are used by millions (Gupta *et al.*, 2015). For example, a caller claiming to be from the victim's bank might say unusual charges were detected on the victim's account, and then ask the victim to confirm their credit card information. Phishers and vishers often use impersonation in their attacks, creating a situation in which the victim feels comfortable or obliged divulging sensitive information (Hadnagy, 2011; The Social Engineering Framework (SEF), 2019).

SE attacks are an increasing global threat (Proofpoint, 2020), with 88% of the organizations receiving targeted phishing attacks and 83% experiencing vishing attacks in 2019. In 2016, 64% of the fraudulent phone calls originated in a country different from the victim, while only 6.6% of legitimate calls originate from international locations (Pindrop, 2017). It is difficult to trace the origination of vishing calls, but Peru, Indonesia, Mexico and India received the most spam calls in 2019, with vishing attacks making up 10-26% of these calls (Kok, 2019). These attacks also have a large impact on the worldwide economy, costing organizations through direct monetary losses, downtime hours and remediation time (Federal Bureau of Investigation (FBI), 2017; Proofpoint, 2020). In just 12 months, 56 million Americans experienced scam calls that cost a total of US$19.7bn (Kok, 2020). Therefore, they have been a focus of attention in the cybersecurity community (CyberEdge Group, 2019). A crucial step in recognizing and mitigating these effects is understanding *how* attackers conduct SE.

Efforts to understand persuasion during social engineering attacks

SE attacks rely on an attacker being persuasive. Thus, researchers have investigated how attackers use persuasion techniques, and several collections of persuasion principles have emerged. For example, Gragg (2003) identified seven psychological triggers, Cialdini (2007) reported six principles of influence and Stajano and Wilson (2011) identified seven principles of general scams.

Examples of collections of persuasion principles contains descriptions of each collection of principles.

Gragg's (2003) seven psychological triggers are as follows:

(1) *Strong affect.* A person in a heightened emotional state (e.g. fearful, excited) is less likely to think reasonably and more likely to be influenced or persuaded.

(2) *Overloading.* When a person is rapidly given too much information, their senses are overloaded, and they are unable to logically evaluate the given arguments. Thus, they become more likely to accept what is being said than they otherwise would be.

(3) *Reciprocation.* People tend to follow the social rule of "returning the favor," repaying social debts to others who (appear to) have helped them in the past.

(4) *Deceptive relationships.* An attacker who establishes a relationship with a victim under false pretenses (e.g. giving the victim information, mentioning a common enemy) can build trust and more easily exploit their victim.

(5) *Diffusion of responsibility and moral duty.* People can be manipulated into feeling that they will not be held solely responsible for their actions, or that their actions are their "moral duty".

(6) *Authority*. People are conditioned to respond to, and not to question, someone who is supposedly in authority.

(7) *Integrity and consistency*. People tend to follow through with what they say they will do and usually believe that others are honest and truthful.

Cialdini's (2007) six principles of influence are as follows:

(1) *Authority*. People tend to not question authority.

(2) *Social proof*. People want to be a part of what others around them are doing, especially if any risks involved are shared by the group.

(3) *Liking/similarity*. People tend to trust and be persuaded by those they know or like, those who are similar to themselves and those they find attractive or credible.

(4) *Commitment/consistency*. People want to be consistent in their actions and feel obligated to honor previously made commitments.

(5) *Scarcity*. People have an emotional response when potential outcomes have limited availability, or there is a restricted amount of time in which they have to respond.

(6) *Reciprocation*. People are obligated by social norms to repay others' actions.

Stajano and Wilson's (2011) seven principles of scams are as follows:

(1) *Distraction*. People are focused solely on what grabs their interest, allowing a scammer to act without being noticed.

(2) *Social compliance (authority)*. People are societally conditioned to suspend suspiciousness of those who appear to be in authority.

(3) *Herd (social proof)*. People believe that there is safety in numbers and let their guard down when risks appear to be shared with those around them.

(4) *Dishonesty*. People can be hooked by dishonestly, and even illegally, participating in a scam, allowing exploitation later on and reducing the likelihood that the victim will go to the authorities.

(5) *Kindness*. People tend to be willing to help others, even volunteering to do so without prompting or reciprocation on the part of the scammer.

(6) *Need and greed (visceral triggers)*. People are driven and distracted by what they need and desire in their current context. Thus, people are less likely to question offers that fulfil their wants and needs.

(7) *Time*. People sacrifice full and proper assessment, reasoning and rationality when under time pressure to make a decision.

These principles have been useful in phishing research. For example, Zielinska *et al.* (2016) examined a set of phishing emails from three US universities using Cialdini's (2007) principles. Similarly, Lawson *et al.* (2017) used a subset of those principles to label emails, aiming to determine technique effectiveness.

An integrated taxonomy of persuasion during social engineering attacks

Ferreira *et al.* (2015) offer a more comprehensive codification of persuasion techniques that integrates previous works and is specific to SE attacks. The researchers integrated principles mentioned in Gragg (2003), Cialdini (2007) and Stajano and Wilson (2011)

(Ferreira *et al.*, 2015, for detailed method). The researchers' final list of five principles of persuasion in SE (PPSEs) is summarized below:

(1) *Authority*. People are conditioned to respond to authority and tend to follow those they think are experts or authority figures.

(2) *Commitment, reciprocation and consistency*. People have more confidence in decisions after publicly committing to following through with a given action. People also tend to believe others, desire to appear consistent in their actions and reciprocate the acts of others.

(3) *Distraction*. People singularly focus attention on their needs, potential gains or losses, time pressure, etc. while ignoring other things that might be happening around them.

(4) *Liking, similarity and deception*. People prefer and listen to others that they know or like, are similar to or familiar with and/or are attracted to.

(5) *Social proof*. People tend to go along with the crowd and want to be included. They feel diminished responsibility for their actions and let their guard down when others appear to be involved in the same behaviors and risks.

To validate this taxonomy, Ferreira and colleagues investigated persuasion techniques in phishing emails. They found several important trends regarding PPSE occurrence and co-occurrence. First, *liking, similarity, and deception*; *authority*; and *distraction* were the three most common principles used in phishing emails, with fewer occurrences of *commitment, reciprocation and consistency* (Ferreira and Chilro, 2017; Ferreira *et al.*, 2015; Ferreira and Lenzini, 2015). Second, the most common PPSE pairs included *authority* and/or *distraction*, suggesting attackers use these PPSEs in conjunction with other PPSEs (Ferreira *et al.*, 2015; Ferreira and Jakobsson, 2016). Finally, *social proof* was rarely used in phishing emails (Ferreira and Lenzini, 2015; Ferreira and Teles, 2019).

An understanding of how persuasion principles are used in SE attacks could be beneficial. First, users could be trained to avoid scams by recognizing attackers' persuasion techniques. Such training is especially relevant for those who manage large amounts of sensitive information, as these individuals are attractive targets. Second, including information about methods today's scammers use could ensure emerging security professionals are aware of real-world threats. Current penetration testers can also benefit from analyzing persuasion techniques to better identify clients' vulnerabilities. Finally, analyzing persuasion principles in SE attacks can identify means for separating them from legitimate solicitations by using different elements than current spam filters.

Present study: social engineering principles in vishing attacks

To date, research has not examined how persuasion principles are used in *vishing* attacks. The use of persuasion principles in vishing and phishing attacks could differ. Vishing involves a continuous verbal interaction between attacker and victim, rather than static text and visual elements. Real-time interaction with the victim could change which persuasion principles an attacker uses. Further, vishing attacks often require the victim to interact and comply with an attacker through multiple steps, which leaves more chances for victims to discover scams compared to just having to click a link or open an attachment.

The present study used 86 examples of real-world *vishing* attacks, which were coded using questions derived from Gragg (2003), Cialdini (2007), Stajano and Wilson (2011) and

Mouton *et al.* (2014) according to Ferreira and colleagues' (2015) five PPSEs. The present study answered the following questions:

Q1. How frequently were the various PPSEs used in the vishing attacks?

Q2. Were certain PPSEs used in the vishing attacks more often than others?

Q3. Which PPSEs were used in the majority of the vishing attacks?

Q4. How frequently did certain PPSEs co-occur in the vishing attacks?

Q5. Were certain sets of PPSEs used more often than others?

Q6. How were those sets of PPSEs implemented (i.e. what specific elements of the attack contributed to the presence of each PPSE)?

Method

Design overview

Real-world examples of *vishing* attack conversations were collected. Each example was coded to determine which persuasion principles were present and how they were implemented. This was similar to how others investigated *phishing* attacks (Ferreira *et al.*, 2015).

Vishing example selection and data set characteristics

We searched Academic Search Complete, YouTube, Google Scholar and Google using the following search terms alone and in combination: *phone, phone scam, social engineering, unsuccessful vishing attempts, vishing, vishing examples*. To be included, the article or video had to contain specifics about a real-world phone interaction between a visher and intended victim. In total, 68 articles and videos were accepted into the study. Some articles and videos contained multiple examples of vishing, each of which was coded separately. Total 86 vishing examples were coded.

Those 86 vishing attacks were a diverse set. The attacker or victim initiated 61 (71%) or 18 (21%) of those attacks. For the remaining seven (8%) attacks, it was unclear who initiated the attack. Information, monetary gain or both were the goal of 39 (45%), 44 (51%) and three (4%) attacks, respectively. Fraud characteristics (Beals *et al.*, 2015) also varied across attacks. For example, individuals or organizations were targeted in 65 (76%) or 21 (24%) of the attacks. Further, the expected benefits or consequences of the fraud concerned consumer products and services, debt collection, personal relationships, prizes or grants or charitable donations in 34 (40%), 25 (29%), 15 (17%), seven (8%) and two (2%) of the attacks, respectively. For the remaining three (3%) attacks, the exact nature of the expected benefits or consequences was unclear.

Coding scheme

A coding scheme was developed to determine the presence and absence of each of the five PPSEs:

(1) authority;

(2) commitment, reciprocation and consistency;

(3) distraction;

(4) liking, similarity and deception; and

(5) social proof.

That scheme consisted of 34 questions, each designed to capture one aspect of persuasion. The questions were developed from persuasion principles presented in the three taxonomies Ferreira *et al.* leveraged (Cialdini, 2007; Gragg, 2003; Stajano and Wilson, 2011) as well as a fourth taxonomy (Mouton *et al.*, 2014). Questions derived from Gragg (2003), Cialdini (2007) and Stajano and Wilson (2011) were divided into Ferreira's five PPSEs according to the guidelines presented in Ferreira *et al.* (2015). Questions derived from Mouton *et al.* (2014) were divided into the five PPSEs based on consensus among the authors. There were between five and nine questions assigned to each principle. The coding scheme is provided in the Appendix.

Coding process, interrater reliability and final code selection

Two researchers coded each of the 86 vishing examples. For a given example, each PPSE was examined by answering the questions assigned to each principle (e.g. *Does the scammer claim to be a member of a reputable institution?* corresponded to *authority*; see Appendix). The questions were answered by the two coders with *yes* if the question was true, *no* if the question was not true or *unable to tell* if there was not enough information present to decide. The *no* and *unable to tell* codes were collapsed into a single *no* code because both concerned the absence of the PPSE. Interrater reliability was assessed using Holley and Guilford's G (Holley and Guilford, 1964; Xu and Lorber, 2014) with a criterion of 0.6, as suggested by Cicchetti (1994). Reliability was sufficiently high between the two coders ($G = 0.64$). One code set was chosen at random to serve as the final codes for analysis.

Our primary interest was whether a PPSE was present or absent. Therefore, after the final code set was chosen, it was determined whether each of the five PPSEs was *present* or *absent* for each example. If the code for one or more questions assigned to a principle was *yes*, the overall principle was coded as *present* for the given example. If the codes for all the questions assigned to a principle were *no*, the overall principle was coded as *absent*. The final *present/absent* codes for each PPSE under each example were used to analyze how frequently PPSEs were used, while the more nuanced question-level codes were used to explore how the PPSEs were implemented.

Results

How frequently were the various principles of persuasion in social engineering used in the vishing attacks?

We determined the prevalence of each individual PPSE in the 86 vishing attack examples. *Authority* was the most commonly used PPSE (95.3% of cases), followed by *social proof* (90.7%); *distraction* (89.5%); and *liking, similarity and deception* (75.6%). *Commitment, reciprocation and consistency* was only observed in 25.6% of cases (Figure 1).

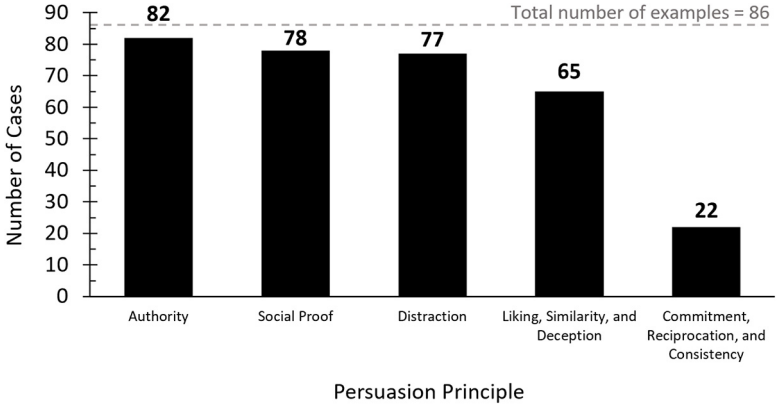
Inspection of Figure 1 suggests vishers:

(1) used certain PPSEs more often than others; and

(2) used certain PPSEs during the majority of attacks. Those possibilities are explored in the following sections.

Were certain principles of persuasion in social engineering used in the vishing attacks more often than others?

We computed the binomial probability (Ott *et al.*, 2016) associated with the observed frequency for a given PPSE, i.e. $P(X)$, setting the probability of the null hypothesis (p) to the proportion of vishing attacks that used the next least frequent PPSE. Thus, we evaluated



Note: The dashed line represents the total number of examples (86)

Figure 1. Overall PPSE prevalence across vishing examples

whether the observed frequency for the target PPSE (X) differed from the frequency that would have been expected based on the next least frequent PPSE given 86 total cases (n):

$$P(X) = \frac{n!}{(n-X)!X!} \times (p)^X \times (1-p)^{n-X} \quad (1a)$$

For example, to determine whether the observed frequency of *authority* differed from what would be expected given the frequency of the next least frequent PPSE, *social proof*, the following calculation would be performed:

$$P(82) = \frac{86!}{(86-82)!82!} \times (.907)^{82} \times (1-.907)^{86-82} \quad (1b)$$

In this and all subsequent analyses, we decided against adjusting the α level to guard against inflation of Type I error (Ott *et al.*, 2016). Instead, we used the per-comparison error, i.e. $\alpha = 0.05$ for each test. We thought this less conservative approach was appropriate given that this study is exploratory and is the first of its kind.

Table 1 presents the associated binomial probabilities. The results suggest *authority*, *social proof* and *distraction* were used equally often. Further, they were used more often than *liking, similarity and deception*. Finally, *liking, similarity and deception* was used more often than *commitment, reciprocation and consistency*.

Which principles of persuasion in social engineering were used in the majority of the vishing attacks?

We computed the binomial probability associated with the observed frequency for a given PPSE, setting the probability of the null hypothesis to 0.75, i.e. our operational definition of "majority." Thus, we evaluated whether the observed frequency for the target PPSE differed from the frequency that would be expected if that PPSE was used in a majority of the attacks (equation (1a)). Binomial probabilities that were significantly above, or did not significantly differ from, "majority" were considered as "majority".

Table 2 presents those binomial probabilities. The observed frequencies for *authority, commitment, reciprocation, and consistency, distraction* and *social proof* differed significantly

from what would be expected if that a PPSE was used in a majority of the vishing attacks. That result, coupled with the absolute values of the observed frequencies, suggests *authority*, *distraction* and *social proof* occurred more often than would be expected if they were used in the majority of attacks, and *commitment, reciprocation and consistency* occurred less often than would be expected if it was used in the majority of attacks. The observed frequency for *liking, similarity and deception* did not differ significantly from what would be expected if that a PPSE was used in a majority of the vishing attacks. Therefore, *authority*, *distraction* and *social proof* were used in the vast majority of vishing attacks; *liking, similarity and deception* was used in the majority of attacks; and *commitment, reciprocation and consistency* was not used in the majority of attacks.

Overall, these results suggest vishers used multiple PPSEs per attack. We next examine which PPSEs were jointly used.

How frequently did certain principles of persuasion in social engineering co-occur in vishing attacks?

We also determined how frequently PPSEs co-occurred. Grouping examples by the PPSEs they contained yielded 14 unique PPSE sets. The PPSE set of *authority*, *distraction*, *liking, similarity, and deception* and *social proof* (ADLS) was the most common (48.2% of cases), followed by *authority*, *commitment, reciprocation, and consistency*, *distraction*, *liking, similarity and deception*, and *social proof* (ACDLS; 14.1%), and *authority*, *distraction* and *social proof* (ADS; 14.1%). Eleven other profiles had three or fewer instances each (Figure 2).

Table 1.
Binomial probabilities associated with the observed frequency for a given PPSE, setting the probability of the null hypothesis to the proportion of vishing attacks that used the next most frequent PPSE

Comparison	Binomial probability
Authority > Social proof?	$p = 0.053$ NS
Social proof > Distraction?	$p = 0.137$ NS
Distraction > Liking, similarity and deception?	$p = 0.0006$
Liking, similarity and deception > Commitment, reciprocation and consistency?	$p < 0.000001$

Notes: Pairs of PPSEs being compared are presented with associated p-values. "NS" denotes probabilities that were not statistically significant. All others were statistically significant

Table 2.
Binomial probabilities associated with the observed frequency for a given PPSE, setting the probability of the null hypothesis to 0.75, i.e. our operational definition of "majority"

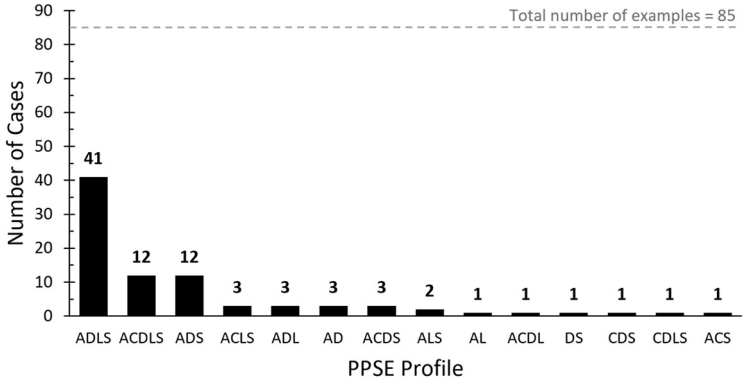
PPSE	Binomial probability
Authority	$p < 0.000001$
Commitment, reciprocation and consistency	$p = 0.0001$
Distraction	$p = 0.0004$
Liking, similarity and deception	$p = 0.1$ NS
Social proof	$p < 0.000001$

Notes: "NS" denotes probabilities that were not statistically significant. All others were statistically significant

Were certain sets of principles of persuasion in social engineering used more often than others? We computed the binomial probability associated with the observed frequency for a given PPSE set, setting the probability of the null hypothesis to the percentage of vishing attacks that used the next least frequent PPSE set. Thus, we evaluated whether the observed frequency for the target PPSE set differed from the frequency that would have been expected based on the next least frequent PPSE set (equations (1a) and (1b)).

Table 3 presents those binomial probabilities. The results indicate ADLS was used more frequently than ACDLS, ACDLS and ADS were used equally often and ADS was used more frequently than ACLS. The remaining comparisons were not significant. Collectively, these results suggest vishers used three PPSE sets – ADLS, ACDLS and ADS – more frequently than other sets of persuasion techniques.

How were those principles of persuasion in social engineering sets implemented? ADLS, ACDLS and ADS occurred more frequently than other PPSE sets (Table 3). Although these sets are comprised of similar PPSEs, how exactly those PPSEs were implemented may



Notes: The dashed line represents the total number of cases (85). One example was excluded from analysis because it only contained one PPSE. *A* = authority; *C* = commitment, reciprocation and consistency; *D* = distraction; *L* = liking, similarity and deception; and S = social proof

Figure 2. Overall prevalence of observed PPSE sets across vishing examples

Table 3. Binomial probabilities associated with the observed frequency for a given PPSE set, setting the probability of the null hypothesis to the percentage of vishing attacks that used the next most frequent PPSE set

Comparison	Binomial probability
ADLS > ACDLS?	$p < 0.000001$
ACDLS > ADS?	n/a (same frequency)
ADS > ACLS?	$p = 0.00004$
Rest of pairs?	$p > 0.05$ NS

Notes: "NS" denotes probabilities that were not statistically significant. All others were statistically significant

vary across sets, which would be lost if PPSE sets were not analyzed separately. Therefore, these three PPSE sets were individually examined, with each question serving as an "element" that contributes to the implementation of a specific PPSE.

For each PPSE set, we computed the binomial probability (equation (1a)) associated with the observed frequency for a given element, setting the probability of the null hypothesis to 0.75. Thus, we evaluated whether the observed frequency for the target element differed from the frequency that would be expected if that element was used in the majority of the attacks for that PPSE set. Binomial probabilities that were significantly above, or did not significantly differ from 0.75, were considered as "majority." This allowed us to determine the characteristics of a typical implementation of each kind of vishing attack.

Authority – distraction – liking, similarity and deception – social proof (ADLS). The binomial probability results indicate five elements occurred at or above the frequency expected if that element was used in a majority of ADLS attacks: S1 (p = 0.000008), A3 (p = 0.009), D7 (p = 0.134), A2 (p = 0.112) and D9 (p = 0.083; Figure 3). All other values were significantly less than "majority" ($0.000000 < p < 0.033$). This suggests ADLS-type attacks involve vishers implying they have authority to access the requested information (A2), claiming to be from a reputable institution (A3), expressing to the victim that there are potential benefits involved if they comply (D7 and S1) and mentioning negative consequences if they do not comply (D9). Interestingly, no elements related to *liking, similarity and deception* occurred in a majority of ADLS-type attacks, which suggests how exactly *liking, similarity and deception* was implemented in ADLS-type attacks varied from attack to attack.

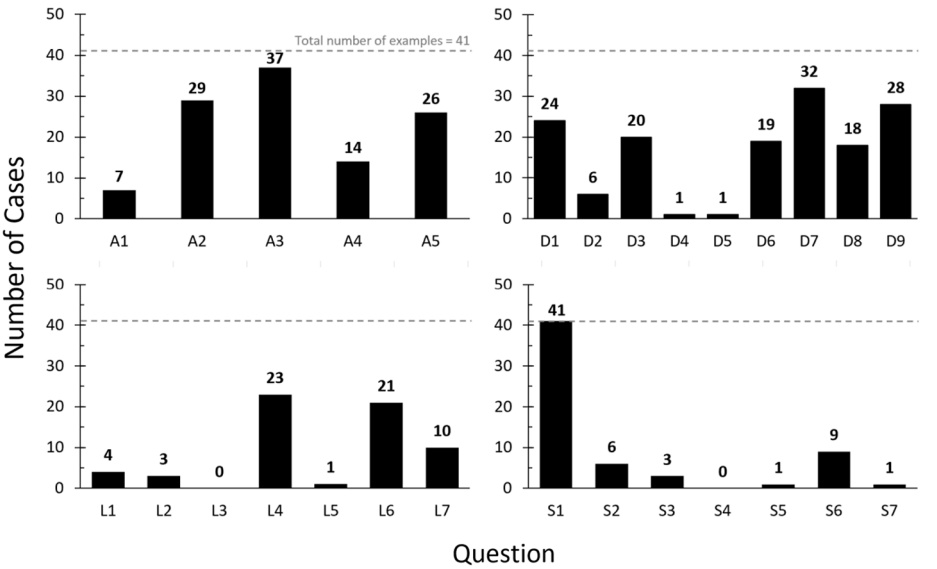


Figure 3.
Overall prevalence of all PPSE elements across ADLS examples

Notes: The dashed lines represent the total number of ADLS examples (41). *A* = authority; *D* = distraction; *L* = liking, similarity and deception; *S* = social proof. Refer to the Appendix for question wording

Authority – commitment, reciprocation and consistency – distraction – liking, similarity and deception – social proof (ACDLS). The binomial probability results indicate eight elements occurred at or above the frequency expected if that element was used in a majority of ACDLS attacks: S1 ($p = 0.031$), A2 ($p = 0.232$), C5 ($p = 0.258$), A3 ($p = 0.194$), S6 ($p = 0.194$), D3 ($p = 0.103$), D6 ($p = 0.103$) and L6 ($p = 0.103$; Figure 4). All other values were significantly less than "majority" ($0.000000 < p < 0.04$). This suggests ACDLS-type attacks involve vishers claiming to be from a reputable institution (A3), claiming to have the authority to access the requested information (A2), emphasizing that the victim is committed to helping them (C5), giving the impression that the requested information is time-sensitive (D3), distracting the victim from thinking about potential consequences (D6) and stressing the benefits (S1) and social correctness (S6) of compliance. These attacks also involve attackers providing some kind of "proof" of their credibility (L6).

Authority – distraction – social proof (ADS). The binomial probability results indicate four elements occurred at or above the frequency expected if that element was used in a

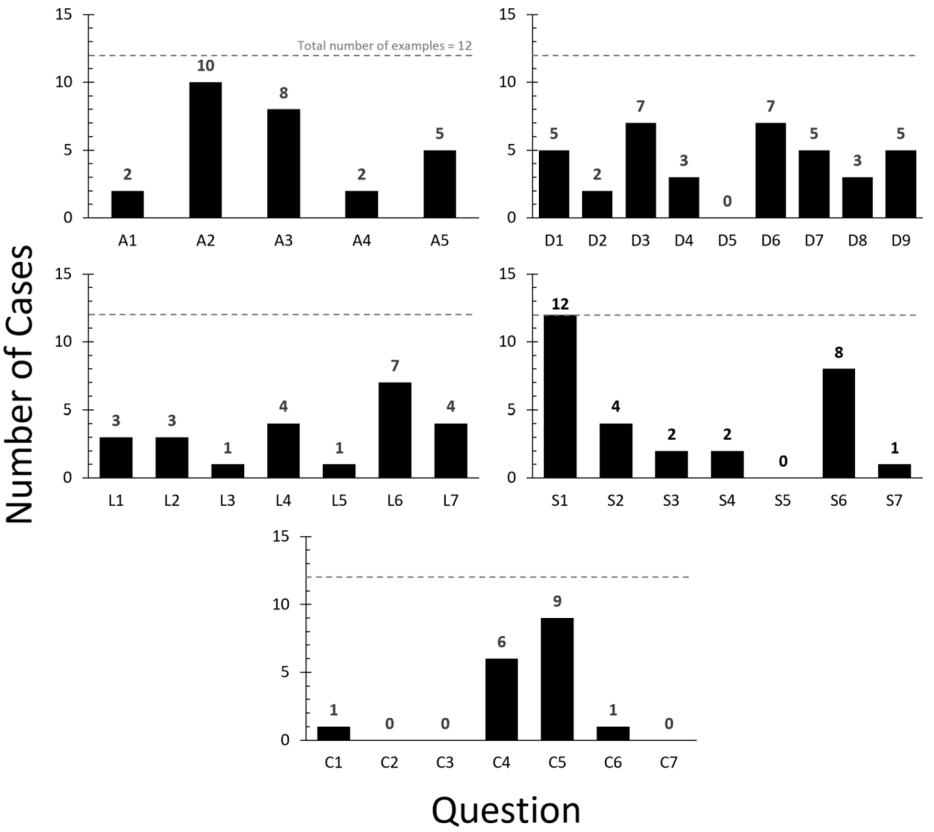


Figure 4.
Overall prevalence of all PPSE elements across ACDLS examples

Notes: The dashed lines represent the total number of ACDLS cases (12). *A = authority; C = commitment, reciprocation and consistency; D = distraction; L = liking, similarity, deception; and S = social proof*

majority of ADS attacks: S1 ($p = 0.032$), A3 ($p = 0.127$), D7 ($p = 0.258$) and D1 ($p = 0.103$; Figure 5). All other values were significantly less than "majority" ($0.000000 < p < 0.04$). This suggests ADS-type attacks involve vishers claiming to be a member of a reputable institution (A3), heightening the victim's emotional state (D1) and expressing the potential benefits of compliance (D7 and S1).

Discussion

This study is the first to investigate how persuasion principles are used during *vishing* attack conversations. We examined five PPSEs (*authority*; *commitment, reciprocation and consistency*; *distraction*; *liking, similarity and deception*; and *social proof*; Ferreira *et al.*, 2015) to evaluate persuasion in vishing attacks. We examined how often the PPSEs were used in 86 vishing attacks, whether some were used more than others and which were used in a majority of the attacks. We also examined the frequency of PPSE co-occurrence, evaluated whether certain PPSE sets were used more than others and identified how those PPSE sets were implemented to provide a detailed picture of different attack types. Our findings indicate that attackers use PPSEs in real-world attacks, verifying the taxonomy of persuasion in SE developed by Ferreira and colleagues.

Vishing

The present results are the first to indicate that *authority (A), social proof* (S) and *distraction (D)* were the most widely used PPSEs, followed by *liking, similarity and deception (L)*. All four of those PPSEs occurred in a majority of vishing attacks, while *commitment, reciprocation and consistency (C)* did not. We also found that certain sets of PPSEs were used more than others, with each set implementing persuasion elements in different ways. In the most prevalent set, ADLS, vishers claim to be from known institutions and aim to convince the victim that there are personal benefits if the victim complies. By contrast, in ACDLS-type attacks, vishers

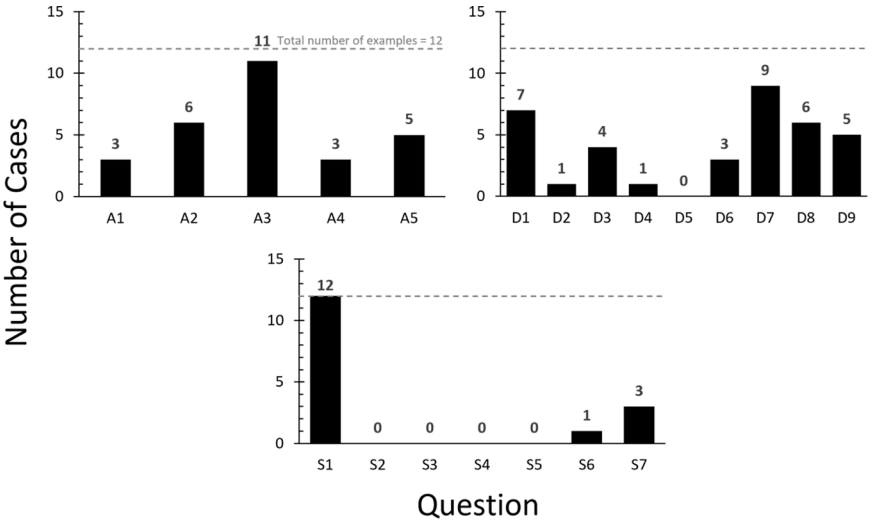


Figure 5.
Overall prevalence of all PPSE elements across ADS examples

Notes: The dashed lines represent the total number of ADS examples (12). *A = authority, D = distraction and S = social proof*

incorporated more persuasion elements overall, especially distraction, and tried to convince the victim that they are committed to helping the visher. ACDLS-type attacks also involved the attacker giving some kind of "proof" that they are legitimate, which is not seen in the majority of ADLS attacks. This may indicate the attacker feels it is necessary to provide credentials only when the vishing attack becomes more complex. In ADS-type attacks, vishers used threats of negative consequences, in addition to describing benefits, and used emotional states to distract their victim. Although implementations varied among these three sets of PPSEs, vishers consistently used two elements in a majority of attacks: claiming to be a member of a reputable institution (A3) and stressing the benefits of compliance (S1). Thus, although there are variable ways a scammer can engineer a vishing attack, these two elements should be paid particular attention in efforts to mitigate vishing attacks. These findings provide a foundation for future research on persuasion during vishing attacks.

When analyzing the prevalence of specific elements that contributed to each PPSE, we noticed that two notable elements were not present in many vishing examples: *Does the scammer claim to have authority over the victim?* and *Does the scammer state or imply that they are in a hurry or otherwise have limited time to converse with the victim?*. One might expect vishers would claim to be authority figures; however, we found instead that vishers attempt to build authority by claiming to be from a reputable institution and suggesting they have authority to access the requested information. This may be because the hierarchy within the victim's own company is quicker and easier for the victim to verify compared to an outside institution. Similarly, we found vishers did not apply time pressure by pretending to be in a hurry, but instead claimed the requested information itself was time-sensitive. The latter tactic is potentially less disconcerting to a victim, allowing the attacker to apply time pressure in a less overt way.

Some elements could be socially strange in a phone call, contributing to their disuse. For example, many elements of *commitment, reciprocation and consistency* (e.g. *Does the scammer perform a kind gesture or a favor toward the victim?*) might seem out of place coming from an out-of-the-blue caller and could alert a victim to the attack. Similarly, many elements of *liking, similarity and deception* involve the attacker making themselves more appealing and similar to the victim (e.g. *Does the scammer make themselves attractive to or flirt with the victim in some way?*). It is unlikely a genuine caller would do so. Thus, vishers may not use these strategies so as to avoid suspicion. Discovering the absence/rarity of certain persuasion elements contributes to a larger understanding of vishing attacks and indicates vishers are socially savvy.

Determining which persuasion principles are used in vishing calls could be useful in training personnel to recognize and appropriately respond to attacks. Many vishing calls are directed at employees who manage customer information, who may not have specific training in recognizing complex over-the-phone attacks compared to other SE attacks, such as phishing. These employees are often the first line of defense for a company in protecting data and could therefore benefit greatly from an informed vishing-specific training program. For example, we found attackers stressed the benefits of complying with their request in a majority of vishing calls. That could be used to help develop a training program that encourages employees to be suspicious of a caller who offers benefits in exchange for sensitive information.

Academic cybersecurity curricula could also benefit from our findings. We examined recent, real-world examples of vishing attacks and therefore provide an up-to-date analysis of persuasion in vishing. New and evolving cybersecurity threats appear regularly, which makes recent SE attack data crucial to providing current and effective training. Determining how scammers implement persuasion principles can also assist penetration testers in more

accurately representing today's attackers and finding vulnerabilities that may not have otherwise been discovered. For example, we found vishers apply time pressure by claiming the requested information is time-sensitive, but not by claiming to be in a hurry. A penetration tester could use this information to realistically role-play a visher.

These findings could also be useful in further expanding SE defenses against vishing attacks. Current call screening is often limited to detecting and alerting users of potential scam calls based on known scam phone numbers. More advanced technologies provide an extra layer of protection against spoofed phone numbers by asking the caller who they are and why they are calling, providing a real-time transcript of the exchange to the user and allowing them to decide what to do with the call. These antispam bots are often effective against robocalls, but vishers can easily give realistic and persuasive responses, convincing the victim to pick up the phone. If patterns of PPSEs are found to be used in vishing calls, or even in short responses to call screening questions, natural language processing methods could be used to identify and alert users to these attacks, even when carried out by a live attacker. For example, the current study found that in one kind of attack (ADS-type attacks), vishers claim to be from a reputable institution, try to heighten the victim's emotional state and express benefits of compliance. If a caller mentions they are from a bank, claims the recipient's account is compromised and they will lose a large amount of money and offers to protect the recipient's assets if they comply, an automated system could detect this combination of elements and alert the call recipient of the potential scam. The potential for drastically improving technologies related to fraud and scam detection warrants further investigation into how attackers use persuasion techniques in more complex, person-to-person SE attacks.

Vishing vs phishing

Research concerning *phishing* attacks revealed *social proof* was not widely used (Ferreira and Lenzini, 2015; Ferreira and Teles, 2019); however, we found *social proof* was highly prevalent in *vishing* attacks, a key finding that reveals a potential difference between SE attack vectors. Specifically, we discovered vishers rely heavily on stressing the benefits of compliance when implementing *social proof*. This same strategy may be less feasible to implement through textual and graphical elements of phishing emails, or less convincing than a person-to-person conversation. Phishing emails are also typically sent to many potential victims, while vishing calls may be directed toward one victim at a time; therefore, they may require more personal persuasion strategies, such as offering benefits in exchange for help from a victim in real time.

These findings demonstrate for the first time that attackers rely heavily on enticing victims with the potential benefits of helping (*social proof*) over the phone when compared to over email. This difference is important because it indicates a fundamental difference between attackers' behavior over email compared to over the phone. Using *social proof* in vishing attacks and not phishing attacks may also be an indicator of attackers' beliefs about the effectiveness of different PPSEs when using one attack vector compared to another.

Theoretical implications

Our research may also have implications for theories of persuasion. The elaboration likelihood model (ELM) describes persuasion in cognitive terms, with individuals using motivation and reasoning when choosing to accept or reject a persuasive message that is delivered through either a direct or peripheral route (Dainton and Zelley, 2005). Our results indicate vishers use peripheral routes to convince someone using emotion, instead of direct

routes that rely on reasoning. Vishers also use some peripheral cues, such as authority and liking, but not others, including commitment and reciprocation.

Future research

It is possible to dive even further into how PPSEs are implemented. For example, what specific language led a coder to respond "yes" to the question *does the scammer provide the victim with some "proof" that they are credible (that they are who they say they are)?* This could give even more insight into these attacks and inform a natural language processing approach to identifying vishing calls.

We examined prevalence of different persuasion principles, but did not look at their effectiveness. It is presumed vishers use these techniques because they are effective, but it may not be the case that every element contributes to an attack's success or failure. Examining which elements drive compliance could further increase understanding of vishing attacks and how to mitigate them. Attack effectiveness may also reveal more implications for the persuasion theory and the effectiveness of different routes to persuasion.

Our data set had relatively few vishing examples in which the victim initiated the call to the attacker. It may be that different PPSEs are used depending on the call initiator, and this possibility should be explored further.

References

Beals, M., DeLiema, M. and Deevy, M. (2015), "Framework for a taxonomy of fraud", available at: http://longevity.stanford.edu/framework-for-a-taxonomy-of-fraud/

Cialdini, R.B. (2007), *Influence: The Psychology of Persuasion*, HarperCollins Publishers, New York, NY.

Cicchetti, D.V. (1994), "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology", *Psychological Assessment*, Vol. 6 No. 4, pp. 284-290, doi: 10.1037/1040-3590.6.4.284.

CyberEdge Group (2019), "2019 Cyberthreat defense report", available at: www.cyber-edge.com/wp-content/uploads/2019/03/CyberEdge-2019-CDR-Report.pdf

Dainton, M. and Zelley, E.D. (2005), *Applying Communication Theory for Professional Life: A Practical Introduction*, SAGE Publications, Inc., Thousand Oaks, CA.

Federal Bureau of Investigation (FBI) (2017), "Business e-mail compromise/e-mail account compromise: the 5 billion dollar scam. Alert no. I-050417-PSA", available at: www.ic3.gov/media/2017/170504.aspx

Ferreira, A. and Jakobsson, M. (2016), "Persuasion in scams", in Jakobsson, M. (Ed.), *Understanding Social Engineering Based Scams*, Springer Science and Business Media, New York, NY, pp. 29-47, doi: 10.1007/978-1-4939-6457-4_4.

Ferreira, A. and Chilro, R. (2017), "What to phish in a subject?", *in International Conference on Financial Cryptography and Data Security, FC 2017 Workshops*, pp. 597-609, doi: 10.1007/978-3-319-70278-0_38.

Ferreira, A. and Teles, S. (2019), "Persuasion: how phishing emails can influence users and bypass security measures", *International Journal of Human-Computer Studies*, Vol. 125, pp. 19-31, doi: 10.1016/j.ijhcs.2018.12.004.

Ferreira, A. and Lenzini, G. (2015), "An analysis of social engineering principles in effective phishing", in *2015 Workshop on Socio-Technical Aspects in Security and Trust*, pp. 9-16, doi: 10.1109/STAST.2015.10.

Ferreira, A., Coventry, L. and Lenzini, G. (2015), "Principles of persuasion in social engineering and their use in phishing", in *Proceedings of the International Conference on Human Aspects of*

Information Security, Privacy, and Trust, HAS 2015, pp. 36-47, doi: 10.1007/978-3-319-20376-8_4.

Gragg, D. (2003), "A multi-level defense against social engineering", available at: www.sans.org/reading-room/whitepapers/engineering/multi-level-defense-social-engineering-920

Gupta, M. and Agrawal, S. (2012), "A survey on social engineering and the art of deception", *International Journal of Innovations in Engineering and Technology*, Vol. 1 No. 1, pp. 31-35.

Gupta, S., Gupta, P., Ahamad, M. and Kumaraguru, P. (2015), "Abusing phone numbers and cross-application features for crafting targeted attacks", arXiv:1512.07330.

Hadnagy, C. (2011), *Social Engineering: The Art of Human Hacking*, Wiley Publishing, Inc., Indianapolis, IN.

Holley, J.W. and Guilford, J.P. (1964), "A note on the G index of agreement", *Educational and Psychological Measurement*, Vol. 24 No. 4, pp. 749-753, doi: 10.1177/001316446402400402.

Kok, K.F. (2019), "Truecaller insights: top 20 countries affected by spam calls and SMS in 2019", available at: https://truecaller.blog/2019/12/03/truecaller-insights-top-20-countries-affected-by-spam-calls-sms-in-2019/

Kok, K.F. (2020), "Truecaller insights 2020 US spam and scam report", available at: https://truecaller.blog/2020/04/16/truecaller-insights-2020-us-spam-scam-report/

Lawson, P., Zielinska, O., Pearson, C. and Mayhorn, C.B. (2017), "Interaction of personality and persuasion tactics in email phishing attacks", in *Proceedings of the Human Factors and Ergonomics Society 2017 Annual Meeting*, pp. 1331-1333, doi: 10.1177/1541931213601815.

Maggi, F. (2010), "Are the con artists back? A preliminary analysis of modern phone frauds", in *Proceedings of the 2010 10th IEEE International Conference on Computer and Information Technology (CIT 2010)*, pp. 824-831, doi: 10.1109/CIT.2010.156.

Mouton, F., Malan, M.M., Leenen, L. and Venter, H.S. (2014), "Social engineering attack framework", in *2014 Information Security for South Africa, IEEE 2014*, pp. 1-9, doi: 10.1109/ISSA.2014.6950510.

Ott, L., Longnecker, M. and Draper, J.D. (2016), *An Introduction to Statistical Methods and Data Analysis*, (7th ed.), Cengage Learning, Boston, MA.

Pindrop (2017), "2017 call center fraud report", available at: www.pindrop.com/resources/download/report/2017-call-center-fraud-report/

Proofpoint (2020), "2020 State of the phish: an in-depth look at user awareness, vulnerability and resilience", available at: www.proofpoint.com/us/resources/threat-reports/state-of-phish

Stajano, F. and Wilson, P. (2011), "Understanding scam victims: seven principles for systems security", *Communications of the Acm*, Vol. 54 No. 3, pp. 70-75, doi: 10.1145/1897852.1897872.

The Social Engineering Framework (SEF) (2019), available at: www.social-engineer.org/framework/general-discussion/

Xu, S. and Lorber, M.F. (2014), "Interrater agreement statistics with skewed data: evaluation of alternatives to Cohen's kappa", *Journal of Consulting and Clinical Psychology*, Vol. 82 No. 6, pp. 1219-1227, doi: 10.1037/a0037489.

Zielinska, O.A., Welk, A.K., Mayhorn, C.B. and Murphy-Hill, E. (2016), "A temporal analysis of persuasion principles in phishing emails", in *Proceedings of the Human Factors and Ergonomics Society 2016 Annual Meeting*, pp. 765-769, doi: 10.1177/1541931213601175.

Appendix

Question	Source(s)
Authority	
A1: Does the scammer claim to have authority over the victim?	[1, 2, 3, 4]
A2: Does the scammer claim to have authority to access the information requested?	[3, 4]
A3: Does the scammer claim to be a member of a reputable institution?	Authors
A4: Does the victim question the authority of the scammer?	[1, 2, 3, 4]
A5: Is it reasonable for the victim to believe that failure to comply with the scammer's request will result in repercussions (e.g. loss of privileges, humiliation, condemnation) based on the scammer's supposed authority?	[2]
Commitment, reciprocation and consistency	
C1: Does the scammer perform a kind gesture or a favor toward the victim?	[1, 2, 4]
C2: Does the scammer perform or claim to have performed a kind gesture toward someone other than the victim?	Authors
C3: Does the scammer try to obligate the victim to reciprocate a kind gesture?	[1, 2, 4]
C4: Does the scammer state or imply that the victim has already committed to helping them (the scammer)?	[1, 2, 4]
C5: Does the scammer state or imply that the victim is committed to helping them based on the victim's job or other obligations?	[1, 2, 4]
C6: Does the scammer state or imply that, based on previous words or actions, it would be inconsistent for the victim to not help the scammer?	[1, 2, 4]
C7: Is it reasonable for the victim to believe that complying with the scammer's request would implicate the victim in activity that is dishonest, illegal or in a legal gray area?	[3]
Distraction	
D1: Does the scammer do anything to heighten the victim's emotional state (e.g. stress, surprise, anger, excitement)?	[1, 2]
D2: Does the scammer give the victim more information than they can process?	[1]
D3: Does the scammer state or imply that the information they are requesting is time-sensitive?	[1, 2, 3]
D4: Does the scammer state or imply that they are in a hurry or otherwise have limited time to converse with the victim?	[1, 2, 3]
D5: Does the scammer state or imply that there is some benefit to complying with their request but that this benefit is of limited quantity?	[2, 4]
D6: Does the scammer attempt to distract the victim from thinking about the intentions or consequences related to the scammer's request?	[3]
D7: Is it reasonable for the victim to believe that if they comply with the scammer's request that they will personally benefit from it?	[3]
D8: Does the scammer state or imply that the consequences of the victim's actions are large?	[3]
D9: Is it reasonable for the victim to believe that if they do not comply with the scammer's request that they will suffer negative consequences because of it?	[3]
Liking, similarity and deception	
L1: Does the scammer establish a relationship with the victim and/or have they established a relationship prior to this phone call?	[1, 4]
L2: Does the victim appear to like the scammer?	[2, 4]
L3: Does the scammer mention any similarities between themselves and the victim?	[2]
L4: Are the scammer and victim of the same gender?	[2]
L5: Does the scammer make themselves attractive to or flirt with the victim in some way?	[2]
	[2, 3]

(*continued*)

Question	Source(s)
L6: Does the scammer provide the victim with some "proof" that they are credible (that they are who they say they are)?	
L7: Is it otherwise reasonable for the victim to believe that the scammer is credible?	[2, 3]
Social proof	
S1: Is it reasonable for the victim to believe that complying with the scammer's request will have benefits (including helping the scammer)?	[1]
S2: Is it reasonable for the victim to believe that they will not be held solely responsible for any negative effects related to complying with the scammer's request?	[1, 2]
S3: Is it reasonable for the victim to believe that any risk associated with helping the scammer is shared by other people as well?	[2]
S4: Does the scammer state or imply that the victim's peers have helped the scammer in this manner in the past?	[3]
S5: Does the scammer state or imply that it is socially correct to help them?	[4]
S6: Is it otherwise reasonable for the victim to believe that it is socially correct to help the scammer?	[4]
S7: Does the scammer state or imply that if the victim does not comply with their request then the victim will be "left out" in some way?	[2]

Sources: [1] Gragg (2003); [2] Cialdini (2007); [3] Stajano and Wilson (2011); and [4] Mouton *et al.* (2014)

Corresponding author
McKenna K. Tornblad can be contacted at: mckenna.tornblad@ttu.edu