How Perceptions of Caller Honesty Vary During Vishing Attacks That Include Highly Sensitive or Seemingly Innocuous Requests

Miriam E. Armstrong, Keith S. Jones, and Akbar Siami Namin
Texas Tech University, Lubbock, Texas, USA

Precis: Participants read written descriptions of vishing conversations. Half included highly sensitive requests; half included seemingly innocuous requests. Participants started in the truth-default state. Several aspects of vishing conversations, including request type, influenced participants' perceptions of visher honesty. It appears that Truth-Default Theory may be a useful tool for understanding vishing attacks.

Running Head: Perceived Honesty During Vishing Calls

Manuscript Type: Research Report
Word Count: 4405

Corresponding Author: Keith S. Jones, keith.s.jones@ttu.edu

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

Abstract

Objective: To understand how aspects of vishing calls (phishing phone calls) influence perceived visher honesty.

Background: Little is understood about how targeted individuals behave during vishing attacks. According to Truth-Default Theory, people assume others are being honest until something triggers their suspicion. We investigated whether that was true during vishing attacks.

Methods: Twenty-four participants read written descriptions of eight real-world vishing calls. Half included highly sensitive requests; the remainder included seemingly innocuous requests. Participants rated visher honesty at multiple points during conversations.

Results: Participants initially perceived vishers to be honest. Honesty ratings decreased before requests occurred. Honesty ratings decreased further in response to highly sensitive requests, but not seemingly innocuous requests. Honesty ratings recovered somewhat, but only after highly sensitive requests.

Conclusions: The present results revealed five important insights: 1) people begin vishing conversations in the truth-default state, 2) certain aspects of vishing conversations serve as triggers, 3) other aspects of vishing conversations do not serve as triggers, 4) in certain situations, people's perceptions of visher honesty improve, and, more generally, 5) Truth-Default Theory may be a useful tool for understanding how targeted individuals behave during vishing attacks.

Application: Those developing systems that help users deal with suspected vishing attacks or penetration testing plans should consider: 1) targeted individuals' truth-bias, 2) the influence of visher demeanor on the likelihood of deception detection, 3) the influence of fabricated situations surrounding vishing requests on the likelihood of deception detection, and 4) targeted individuals' lack of concern about seemingly innocuous requests.

Keywords

vishing; telephone fraud; social engineering; deception detection; Truth-Default Theory

Introduction

Social engineers acquire information by "hacking" humans (Hadnagy, 2010). One form of social engineering is vishing (a portmanteau of "voice" and "phishing"), which occurs when a social engineer attempts to persuade a targeted individual to divulge information such as a username and password, or other employees' names and roles, during a phone call (Griffin & Rackley, 2008; Ollmann, 2007). The visher may use such information for immediate financial gain, or to deceive others during future vishing attacks.

Vishing has grown into a constant and expensive problem (Bullée, Montoya, Pieters, Junger, & Hartel, 2018; Maggi, 2010). Globally, at least 40% of working adults can be expected to experience vishing attacks within a given year (Proofpoint, 2019), and 69% of frauds reported to the Federal Trade Commission occurred over the phone (FTC, 2019). Vishing and other social engineering attacks impact hundreds of thousands of people globally, and approximately 25% of the financial losses stemming from these attacks are never recovered (FBI & IC3, 2019).

To thwart vishing attacks, call blocking technologies have been implemented. However, vishers can spoof caller ID information, rendering such technologies ineffective (Pandit, Perdisci, Ahamad, & Gupta, 2018); thus, individuals continue to receive and respond to vishing calls (Tu, Doupé, Zhao, & Ahn, 2019). When that happens, targeted individuals may benefit from tools, such as a digital assistant, that help them deal with the suspected vishing attack. To develop such tools, we need to understand how targeted individuals experience vishing calls and detect vishers' deception. To date, little empirical research has investigated that topic, however, deception detection theories, such as Truth-Default Theory (TDT) (Levine, 2014b), may prove useful.

Truth-Default Theory

According to TDT, people generally assume others are being honest (Levine, 2014b). That is, people are typically in a truth-default state. This is adaptive because the majority

of day-to-day communications are truthful (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996; Serota, Levine, & Boster, 2010). Thus, it is usually accurate to believe that one's conversation partner is honest (Park & Levine, 2001; Street, 2015). Because of their truth-default state, individuals do not actively evaluate the veracity of what they are being told unless something triggers them to do so (Clare & Levine, 2019; Levine, 2014b).

According to TDT, several triggers are possible (Levine, 2014b). Examples include: 1) the target becoming aware of a motive for lying, 2) the liar saying or doing things associated with a dishonest demeanor, 3) the liar saying or doing things that do not agree with one another, 4) the liar saying or doing things that do not agree with known reality, or 5) a third-party's warning of potential deception (Levine, 2014b). Research regarding triggers is on-going, so this list is likely not exhaustive (Levine, 2014b; Street, 2015).

When a trigger prompts one to leave the truth-default state, one searches for evidence that the suspected message is deceptive. For example, one might question the person with whom they are speaking so as to catch them in a lie (Levine, 2014a). If one acquires sufficient evidence of lying, they will decide the person with whom they are speaking is being deceptive; otherwise, the individual will decide the other person is being honest, and return to the truth-default state (Levine, 2014b).

Vishing Through a Truth-Default Theory Lens

In the context of vishing, TDT suggests targeted individuals will be in the truth-default state at the beginning of the phone conversation. That tendency, which ordinarily is adaptive, now puts them at risk because they will not actively evaluate the veracity of what the visher is telling them unless a trigger prompts them to do so.

What aspects of a vishing call might be triggers? One possibility is the visher's request. Some requests concern highly sensitive information, e.g., personal identification numbers (PINs) or social security numbers (SSNs) (Ollmann, 2007). Other requests concern seemingly innocuous information, e.g., a co-worker's name or vacation schedule, or

job-specific terminology, which can be used to build credibility during future social engineering attacks (Mitnick & Simon, 2011). For example, a social engineer who learned that banks communicate their "merchant identification number" to consumer credit reporting agencies can use that terminology to appear more credible when later vishing someone from such agencies (Mitnick & Simon, 2011). Targeted individuals may be triggered by either request type. After all, both request types involve someone the targeted individual does not know asking them to do something that is at least a little out of the ordinary. With that said, targeted individuals will probably be triggered to a greater degree by highly sensitive requests than seemingly innocuous requests because people are generally unwilling to share highly sensitive information (Phelps, Nowak, & Ferrell, 2000), and may even consider the malicious reasons why the visher requested the sensitive information, which could greatly increase the likelihood that the targeted individual will detect the visher's deception (Bond Jr, Howard, Hutchison, & Masip, 2013).

The Current Experiment

Participants read written descriptions of phone conversations; each described a visher calling a targeted individual and requesting they provide some information or perform some action. In half of the descriptions, vishers requested highly sensitive information; in the other half, vishers requested seemingly innocuous information. Participants rated visher honesty at multiple points during the conversation descriptions.

We made three hypotheses. First, we hypothesized participants would be in the truth-default state when they began reading the conversation descriptions (Hypothesis 1). Thus, we predicted participants would rate vishers as being significantly more honest than neutral at the beginning of the conversation descriptions. Second, we hypothesized that vishing requests would be triggers (Hypothesis 2). Therefore, we predicted honesty ratings would be greater before vishing requests than during requests. Third, we hypothesized that some types of requests would serve as stronger triggers than others (Hypothesis 3).

Specifically, we predicted requests for highly sensitive information, such as PINs and passwords, would lead participants to reduce their honesty ratings to a greater degree than requests for seemingly innocuous information, such as requests for names or for terminology. To our knowledge, this experiment is one of the first to explicitly test hypotheses derived from TDT within a social engineering context.

Method

Participants

Twenty-four participants (17 women; $M_{\text{age}} = 26.36$, $SD_{\text{age}} = 8.36$) were recruited through the university's announcement system and fliers posted on campus. Participants were fluent in English and received $30 for participation. No participants reported taking a cybersecurity course, having cybersecurity work experience, or receiving training about what to do during a vishing call. Three participants reported being vishing attack victims, and one of those participants also reported receiving training on determining whether a phone call was a vishing call. Those participants were not excluded from the sample. Excluding those participants from the data set did not alter our conclusions.

Materials

Examples of successful real-world vishing attacks were identified (Get Safe Online, 2015a, 2015b, 2015c; MadeInSyr, 2013; Mitnick & Simon, 2011; Power & Forte, 2006). Examples had to meet several criteria: the conversation 1) had two speakers, the visher and the targeted individual, 2) began with the visher calling the targeted individual, and 3) was a complete conversation.

From those examples, eight written descriptions of vishing calls were created. All descriptions concerned vishing calls because our hypotheses did not concern differences between vishing and non-vishing calls. Descriptions were written, rather than recordings played for participants, because a recording of a role-played interaction might not

accurately represent the many personality, emotional, and behavioral characteristics that make one a convincing liar (Vrij, Granhag, & Mann, 2010). Descriptions were written, rather than actual interactions with participants, because the latter would make it impossible to control conversation content or the role-played visher's behavior. Brief descriptions of each conversation can be found in Table 1.

Each conversation was divided into 9-30 sections. Sections were 1-2 sentences long and had one speaker. Conversations contained one of two request types: highly sensitive and seemingly innocuous. During highly sensitive requests, vishers asked for information such as login information or a PIN. During seemingly innocuous requests, vishers asked for information such as job-specific terminology or a coworker's name, or asked the targeted individual to perform a task, such as accessing a Web site. Conversations that included highly sensitive or seemingly innocuous requests were similar in the following ways: total number of sections within a conversation, $t(6) = 0.53$, $p = .612$, the portion of the conversation in which the visher was talking, $t(6) = 0.92$, $p = .394$, and the number of sections between the beginning of a conversation and when the visher made their request, $t(6) = 0.43$, $p = .681$.

Perceived honesty ratings were collected with a scale used in previous deception detection research (McCornack, Levine, Solowczuk, Torres, & Campbell, 1992) ($\alpha = .99$). The scale contains four semantic difference items: misleading/not misleading, deceitful/truthful, dishonest/honest, and deceptive/not deceptive. Each item was rated from 1 to 6 (e.g., misleading = 1; not misleading = 6). Those four items were averaged to determine a honesty score between 1 and 6 (6 = greatest level of perceived honesty).

Procedure

The research complied with the APA Code of Ethics, and was approved by the Texas Tech Institutional Review Board. Each participant provided informed consent.

During instructions, participants were told that some speakers might be deceitful.

Table 1
Brief Descriptions of Phone Conversation Stimuli

Conversation Description	Request
Highly Sensitive Request Type	
Conversation 1. The target is the supervisor of a phone company's help desk. The visher claims that the systems are down and asks for the supervisor's username and password so that they may access her account and determine what's going on.	username, password
Conversation 2. The visher calls a small business claiming to be from a bank and concerned about a series of suspicious payments made by the business. The visher requests the target's username and PIN in order to cancel the payments.	username, PIN
Conversation 3. The visher calls a small business claiming to be from a bank and concerned about a series of suspicious debits made by the business. The visher requests the target's PIN in order to cancel the debits.	PIN
Conversation 4. A visher calls a bank claiming to be an employee at another branch of the bank and requesting a customer's signature card which contains sensitive information such as the customer's SSN.	signature card
Seemingly Innocuous Request Type	
Conversation 5. A visher calls a company's tech support team and claims that they can't access a Web site. When the target visits the Web address, the visher is able to access the company's network.	visit Web site
Conversation 6. A visher asks a bank employee to confirm a banking term and claims they wish to know because they are writing a book and want to use the correct terminology.	terminology
Conversation 7. The visher asks a receptionist for the name of an accounting department manager so that they can address an invoice to a specific person.	name
Conversation 8. The visher calls a small business claiming to be from a bank and concerned about a series of suspicious deposits made by the business. The visher asks for the names of business associates who should be notified about the deposits.	name

That was consistent with previous experiments (Granhag & Strömwall, 2001; Levine, Kim, & Blair, 2010; Schindler & Reinhard, 2015), but may have primed participants to perceive dishonesty more readily than they would have otherwise (Clare & Levine, 2019; Parsons, McCormac, Pattinson, Butavicius, & Jerram, 2015).

Participants then completed a practice block, during which they read a written description of a phone conversation that was divided into four sections. Participants read a given section, rated the honesty of the speaker in that section, and then continued to the next section. Participants then completed eight experimental blocks. The order of which was randomized for each participant. Experimental blocks were identical to the practice block except the length of the phone conversation description varied. Finally, participants answered two sets of questions. The first set concerned demographics. The second asked whether participants had taken a cybersecurity course, had cybersecurity work experience, had been trained to recognize vishing calls, had been trained to respond to vishing calls, or had been a vishing victim.

As noted, participants read written descriptions of phone conversations; they did not participate in those conversations. That may have further lessened participants' honesty ratings for several reasons. First, those who observe liars conversing with others perceive liars to be less honest than those who talk with liars (Feeley & DeTurck, 1997). Second, those who do not participate in a conversation experience less mental workload, and thus have more resources to devote to deception detection, than those who participate (Waugh et al., 2000). Third, those who do not participate in a conversation may be somewhat inoculated from the many personality, emotional, and behavioral characteristics that make one a convincing liar (Vrij et al., 2010). Accordingly, perceived honesty in real-world settings will likely be greater than that reported here.

Results

Computing Honesty Ratings Per Conversation Section

Honesty ratings were averaged across participants at four conversation sections: the Beginning, Before Request, Request, and End sections (Table 2). Beginning sections were the first sections in which vishers spoke. Before Request sections were the last sections before requests in which vishers spoke. Request sections were those in which vishers made their requests. End sections were the last sections in which vishers spoke. Mean honesty ratings for each conversation at each section can be found in Table 2.

All conversations contained Beginning and Request sections. For those sections, honesty ratings were averaged across all four conversations within a conversation type. Because stimuli were created based on real-world vishing conversations, not all conversations contained Before Request and End sections. For those sections, honesty ratings were averaged across all conversations that contained those sections within a conversation type (see Table 2 for details). This was acceptable because it did not impair our ability to test our hypotheses.

Analytic Approach

The following sub-sections describe tests performed to investigate three hypotheses and seven research questions related to unexpected trends. We differentiate tests of hypotheses from tests of non-hypotheses by referring to the latter as exploratory. Our narrative interweaves the two test types so the Results section follows the flow of the conversations.

For each test, we computed *t*-tests. Parametric tests were considered appropriate because 1) composite scores derived from sets of response items with discrete values often exhibit the characteristics of an interval scale (Carifio & Perla, 2007), and 2) even if our honesty scores did not exhibit those qualities, *t*-tests are quite robust against violations of the interval level data assumption (Carifio & Perla, 2007; Havlicek & Peterson, 1974).

The *t*-tests compared mean honesty ratings either across conversation sections or

Table 2
Mean (SD) honesty ratings of each conversation during four sections of the phone conversation stimuli

	Conversation Section: Mean (SD)			
Conversation	Beginning	Before Request	Request	End
Highly Sensitive Request Type				
Conversation 1	3.81 (1.52)	1.98 (1.45)	1.21 (0.51)	1.89 (1.27)
Conversation 2	5.17 (1.21)	3.28 (1.71)	2.71 (1.52)	3.63 (1.74)
Conversation 3	5.67 (0.82)	2.09 (1.36)	1.67 (0.84)	–
Conversation 4	5.66 (0.76)	–	4.18 (1.67)	3.94 (1.66)
Across Conversations	5.08 (0.73)	2.45 (1.14)	2.44 (0.71)	3.15 (1.18)
Seemingly Innocuous Request Type				
Conversation 5	5.78 (0.51)	5.47 (0.87)	5.69 (0.62)	5.13 (1.70)
Conversation 6	5.71 (0.75)	4.71 (1.44)	4.74 (1.45)	5.14 (1.42)
Conversation 7	5.14 (1.30)	–	4.89 (1.62)	5.45 (1.26)
Conversation 8	5.50 (0.88)	5.06 (1.46)	4.83 (1.61)	4.69 (1.53)
Across Conversations	5.53 (0.63)	5.08 (0.89)	5.04 (0.86)	5.10 (0.96)

Note: Beginning sections were the first sections in which vishers spoke. Before Request sections were the last sections before the request in which vishers spoke. Request sections were those in which vishers made their requests. End sections were the last sections in which vishers spoke. Honesty scores could range from 1 to 6 (6 = greatest level of perceived honesty).

against the honesty scale's neutral point, i.e., 3.5. For the latter, mean honesty ratings that were significantly greater than 3.5 (neutral) were interpreted as reflecting a truth-default state. Effect sizes were variations of Cohen's d: d_{rm} for paired samples t-tests and d_z for single sample t-tests (Lakens, 2013).

We considered t-tests related to each hypothesis or research question to be a family. We applied a Bonferroni correction so that $\alpha = .05$ for each family (Tabachnick, Fidell, & Ullman, 2007).

Were Participants in the Truth-Default State at the Beginning of Conversations?

We hypothesized participants would be in the truth-default state during Beginning sections (Hypothesis 1). To test that, we conducted two single-sample t-tests ($\alpha = .05/2 = .025$), one for conversations that eventually involved requests for highly sensitive information and another for conversations that eventually involved requests for seemingly innocuous information. Each compared Beginning section honesty ratings to 3.5.

The mean honesty rating for Beginning sections was significantly greater than 3.5 for conversations that eventually involved highly sensitive requests ($M = 5.08$, $SD = 0.73$), $t(23) = 10.53$, $p < .001$, $d_z = 2.15$, as were those for conversations that eventually involved seemingly innocuous requests ($M = 5.53$, $SD = 0.63$), $t(23) = 15.83$, $p < .001$, $d_z = 3.23$. Accordingly, participants were in the truth-default state during Beginning sections, which supports Hypothesis 1.

Were Honesty Ratings Equal at the Beginning of Conversations?

We conducted a paired-samples t-test to explore whether Beginning section honesty ratings were similar across conversation types ($\alpha = .05$). To our surprise, the mean honesty rating for conversations that eventually included highly sensitive requests was significantly less than that for conversations that eventually included seemingly innocuous requests, $t(23) = 4.31$, $p < .001$, $d_{rm} = 0.66$. Accordingly, the visher's initial interaction with their target led participants to perceive vishers who later made highly sensitive requests to be less honest than vishers who later made seemingly innocuous requests.

Why Were Honesty Ratings Unequal Between Conversation Types?

We explored the data at the conversation level to investigate why Beginning section honesty ratings differed between conversation types. Doing so revealed that three of the four conversations that eventually included highly sensitive requests had mean honesty

ratings ($Ms = 5.17 - 5.67$) that were similar to those for the four conversations that eventually included seemingly innocuous requests ($Ms = 5.14 - 5.78$); see Table 2. In contrast, the fourth conversation that eventually included a highly sensitive request, Conversation 1, had a mean honesty rating ($M = 3.81$) that was less than the others. Thus, it appeared Conversation 1 drove the significant difference between conversation types.

To evaluate that, we conducted a paired-samples t-test ($\alpha = .05$), which compared Beginning section mean honesty ratings across conversation types, but this time with Conversation 1 excluded. That test revealed the mean Beginning section honesty rating for conversations that eventually included highly sensitive requests ($M = 5.50$, $SD = 0.75$) was not significantly different from that for conversations that eventually included seemingly innocuous requests, $t(23) = 0.31$, $p = .762$, $d_{rm} = 0.05$, which suggests the difference in Beginning section honesty ratings between the two conversation types was driven by unique characteristics of Conversation 1.

Vishers' opening lines in Conversation 2-8 are conventional; for instance, the visher greets the targeted individual. In contrast, the visher in Conversation 1 begins the conversation by asking the targeted individual if they are having a bad day. That unconventional opening line may have caused participants to perceive the visher as unpleasant or unfriendly. Pleasantness and friendliness are characteristics of an honest demeanor (Levine et al., 2011). Therefore, that unconventional opening line may have served as a trigger because participants may have perceived the visher as having a dishonest demeanor. A dishonest demeanor, while not actually diagnostic of dishonesty, is often perceived as an indication of deception (Levine et al., 2011).

Did Honesty Ratings Decline Before Vishing Requests?

To explore whether honesty ratings declined before vishing requests, we conducted two paired-samples t-tests, one for each conversation type ($\alpha = .05/2 = .025$). Beginning section honesty ratings were compared to those for Before Request sections.

Honesty ratings significantly decreased between the Beginning ($M = 4.88$, $SD = 0.83$) and Before Request ($M = 2.45$, $SD = 1.14$) sections for conversations that eventually included highly sensitive requests, $t(23) = 9.96$, $p < .001$, $d_{rm} = 2.42$. Similarly, honesty ratings significantly decreased between the Beginning ($M = 5.66$, $SD = 0.89$) and Before Request ($M = 5.08$, $SD = 0.62$) sections for conversations that eventually included seemingly innocuous requests, $t(23) = 3.24$, $p = .004$, $d_{rm} = 0.75$. Notably, the former's effect size was three times greater than the latter's. Collectively, these results indicate participants perceived vishers to be less honest than they had at the beginning of the conversation, even before any vishing requests were made, and especially for vishers who later made highly sensitive requests.

Prior to requests, conversations that led to seemingly innocuous requests mostly concerned fairly typical situations, e.g., technical support helping someone to access a Web site (Conversation 5) or a bank employee answering a question about whether their bank uses a given credit reporting agency (Conversation 6). In contrast, prior to requests, conversations that led to highly sensitive requests concerned fairly atypical situations, e.g., the targeted individual being told that their system went down without their awareness (Conversation 1), or dealing with potentially fraudulent transactions (Conversation 2 and 3). Individuals who access a given platform (e.g., Facebook) more frequently and for longer durations are less susceptible to social engineering through that platform than others, presumably because the former are better able than the latter to identify atypical situations, which often signal attacks (Heartfield, Loukas, & Gan, 2016). If that is true for vishing, then the fabricated contexts surrounding highly sensitive requests may have served as a stronger trigger than those surrounding seemingly innocuous requests because participants' suspicions might be proportional to how typical or atypical are those contexts.

Were Participants Still in the Truth-Default State Immediately Before Vishing Requests?

To explore whether participants remained in the truth-default state immediately before the vishing requests, we conducted two single-sample t-tests, one for each conversation type ($\alpha = .05/2 = .025$). Each compared mean Before Request section honesty ratings to 3.5.

Mean Before Request section honesty ratings during conversations that eventually included highly sensitive requests ($M = 2.45$, $SD = 1.14$) were significantly less than 3.5, $t(23) = 4.52$, $p < .001$, $d_z = 0.92$. In contrast, mean Before Request section honesty ratings for conversations that eventually included seemingly innocuous requests were significantly greater than 3.5 ($M = 5.08$, $SD = 0.89$), $t(23) = 8.67$, $p < .001$, $d_z = 1.77$. Thus, participants were not in the truth-default state at this point during conversations that eventually included highly sensitive requests and participants remained in the truth-default state at this point during conversations that eventually included seemingly innocuous requests.

Did Honesty Ratings Decline Further Because of Vishing Requests?

We hypothesized vishing requests would serve as triggers (Hypothesis 2). To evaluate that, we conducted two paired-samples t-tests, one for each conversation type ($\alpha = .05/2 = .025$). Each compared Before Request and Request section honesty ratings.

When vishers made highly sensitive requests, the mean Before Request section honesty rating ($M = 2.45$, $SD = 1.14$) was significantly greater than that for the Request section ($M = 1.86$, $SD = 0.76$), $t(23) = 3.49$, $p = .002$, $d_{rm} = 0.56$. In contrast, when vishers made seemingly innocuous requests, the mean Before Request section honesty rating ($M = 5.08$, $SD = 0.89$) was not significantly different than that for the Request section ($M = 5.09$, $SD = 0.75$), $t(23) = 0.07$, $p = .944$, $d_{rm} = 0.01$. Thus, requests for highly sensitive information served as triggers whereas requests for seemingly innocuous information did not, which partially supports Hypothesis 2 and fully supports Hypothesis 3.

Were Participants in the Truth-Default State When Vishing Requests Were Made?

To explore whether participants were in the truth-default state during Request sections, we conducted two single-sample t-tests, one for each conversation type ($\alpha = .05/2 = .025$). Each compared Request section honesty ratings against 3.5.

When highly sensitive requests were made, honesty ratings were significantly below neutral ($M = 2.44$, $SD = 0.71$), $t(23) = 7.33$, $p < .001$, $d_z = 1.50$. When seemingly innocuous requests were made, honesty ratings were significantly above neutral ($M = 5.04$, $SD = 0.86$), $t(23) = 8.77$, $p < .001$, $d_z = 1.79$. Thus, participants were not in the truth-default state during highly sensitive requests, and were in the truth-default state during seemingly innocuous requests.

Did Honesty Ratings Recover After Vishing Requests?

Typically, vishing requests were not the end of conversations, so it was possible that honesty ratings could have recovered after requests were made. To explore that, we conducted two paired-samples t-tests ($\alpha = .05/2 = .025$), one for each conversation type. Each compared Request section honesty ratings to those for End sections.

For conversations that included highly sensitive requests, the mean Request section honesty rating ($M = 2.70$, $SD = 0.78$) was less than that for the End section ($M = 3.15$, $SD = 1.18$) but the difference was not significant after Bonferroni correction, $t(23) = 2.25$, $p = .034$, $d_{rm} = 0.43$. For conversations that included seemingly innocuous requests, the mean Request section honesty rating ($M = 5.04$, $SD = 0.86$) was not significantly different from that for the End section ($M = 5.10$, $SD = 0.96$), $t(23) = 0.43$, $p = .669$, $d_{rm} = 0.07$. These results suggest honesty ratings did not recover after vishing requests. However, the effect size for conversations that included highly sensitive requests ($d_{rm} = 0.43$) suggests those honesty ratings may have recovered somewhat, although not enough to be statistically significant once the Bonferroni correction was applied.

Were Participants in the Truth-Default State at the End of Conversations?

To explore whether participants were in the truth-default state during End sections, we conducted two single-sample t-tests, one for each conversation type ($\alpha = .05/2 = .025$). Each compared End section honesty ratings to 3.5.

For conversations that included highly sensitive requests, the mean End section honesty rating was not significantly different from 3.5 ($M = 3.15$, $SD = 1.18$), $t(23) = 1.45$, $p = .161$, $d_z = 0.30$. For conversations that included seemingly innocuous requests, the mean End section honesty rating was significantly greater than 3.5 ($M = 5.10$, $SD = 0.96$), $t(23) = 8.17$, $p < .001$, $d_z = 1.67$. Thus, participants were not in the truth-default state at the end of highly sensitive conversations, and were in the truth-default state at the end of seemingly innocuous conversations.

For conversations that included highly sensitive requests, the mean Request section honesty rating was significantly less than 3.5 ($M = 2.44$, $SD = 0.71$) whereas the mean End section honesty rating was not ($M = 3.15$, $SD = 1.19$). That corroborates our earlier argument that honesty ratings for conversations that included highly sensitive requests recovered somewhat by the end of those conversations. What might have driven that recovery is explored in the Discussion section.

Discussion

For vishing conversations that included requests for highly sensitive information, participants entered the conversation in the truth-default state. As those conversations progressed, participants left the truth-default state. Vishers were perceived to be dishonest immediately before vishing requests, and honesty ratings declined further once requests were made. By the end of those conversations, honesty ratings had recovered somewhat, enough that vishers were perceived as neither honest nor dishonest (neutral) rather than dishonest.

For vishing conversations that included requests for seemingly innocuous information,

participants entered the conversation in the truth-default state. As those conversations progressed, honesty ratings declined but not enough for participants to leave the truth-default state. Honesty ratings did not decline further when seemingly innocuous requests were made, and did not recover by the end of those conversations.

Insights Derived From The Present Results

The present results revealed five important insights: 1) people begin vishing conversations in the truth-default state, 2) certain aspects of vishing conversations serve as triggers, 3) other aspects of vishing conversations do not serve as triggers, 4) in certain situations, people's perceptions of visher honesty improve, and, more generally, 5) TDT may be a useful tool for understanding how targeted individuals behave during vishing attacks. Each will be detailed below.

People Begin Vishing Conversations in the Truth-Default State. During Beginning sections, honesty ratings were significantly greater than the honesty scale's neutral value for both conversation types. Thus, our results supported Hypothesis 1, which stated participants would be in the truth-default state at the beginning of vishing conversations. As noted earlier, people often exhibit a truth-bias (Levine, 2019), however, that is not always the case. For example, people exhibit a lie-bias when evaluating online news (Baryshevtsev et al., 2020). Accordingly, it is important to establish that people exhibit a truth-bias at the beginning of vishing conversations.

Certain Aspects of Vishing Conversations Serve as Triggers. Honesty ratings for Conversation 1 during the Beginning section were less than those for the other conversations. As noted earlier, the unconventional opening line in Conversation 1 may have served as a trigger because participants may have perceived the visher to have a dishonest demeanor, which is known to trigger suspicion (Levine, 2014b).

Honesty ratings decreased prior to requests, especially during conversations that led to highly sensitive requests. As noted previously, the relatively atypical contexts that

surrounded highly sensitive requests (e.g., computer systems being down without the targeted individual being aware of that) may have served as a trigger because participants' suspicions may be a function of typicality (Heartfield et al., 2016).

Honesty ratings also decreased as a result of highly sensitive requests. During such requests, vishers asked targeted individuals to provide personal information such as usernames (Conversation 1 and 2), passwords (Conversation 1), or PINs (Conversation 2 and 3), or to send vishers a document that contained such information (Conversation 4). Most people have been explicitly told that they should not share such information with others. Therefore, requesting that information may have served as a trigger because the visher violated what participants were taught or because the request caused participants to consider the vishers' motives for requesting that information. Requests for highly sensitive information also served as triggered during phishing attacks (Downs, Holbrook, & Cranor, 2006; Furnell, 2007), which suggests such requests serve as triggers across social engineering attack types.

In summary, these observations collectively shed light on aspects of vishing conversations that may serve as triggers. Specifically, aspects of vishing conversations will likely encourage people to leave the truth-default state when they: 1) cause the visher to be perceived as unfriendly or unpleasant, 2) cause the situation to be perceived as atypical, or 3) violate what participants were taught.

Certain Other Aspects of Vishing Conversations Do Not Serve as Triggers. Honesty ratings did not decrease as a result of seemingly innocuous requests, and were significantly greater than neutral during those requests. During such requests, vishers asked targeted individuals to visit a Web site (Conversation 5), answer a question about terminology (Conversation 6), or provide a co-worker's name (Conversation 7 and 8). Those requests may not have served as triggers because participants may not have been warned about such attacks or understood how those requests could be used for nefarious purposes. Regarding the latter, some users are unaware that phishers sometimes request

seemingly innocuous information (Greene, Steves, Theofanos, & Kostick, 2018); the same may be true for vishing attacks. Alternatively, those requests may not have served as triggers because such requests are commonplace but rarely attacks, which makes them very difficult to detect when they are attacks (Sawyer & Hancock, 2018). Given that participants stayed in the truth-default state during seemingly innocuous requests, people will likely not detect deception when vishers make requests that are not typically mentioned during social engineering training, for which a nefarious motive is not obvious, or are only rarely associated with attacks.

In Certain Situations, Perceptions of Visher Honesty Improve. Honesty ratings increased between Request and End sections, but only after highly sensitive requests. Vishers who made seemingly innocuous requests mostly made polite conversation after making their requests. In contrast, vishers who made highly sensitive requests often provided follow-up information that made their requests seem reasonable. For example, vishers explained why a problem was occurring (Conversation 1), explained why it would be necessary to generate a passcode (Conversation 2), or provided a security code (Conversation 4). A similar approach is used when crafting spear-phishing emails, making them very difficult to differentiate from legitimate emails (Butavicius, Parsons, Pattinson, & McCormac, 2016). Providing plausible explanations is a characteristic of an honest demeanor (Levine et al., 2011). Therefore, providing follow-up information that made highly sensitive requests seem plausible may have partly assuaged participants' concerns, leading participants to perceive those vishers as more honest than they were perceived to be before.

Truth-Default Theory May Be A Useful Tool. According to TDT, people generally assume others are being honest, unless something triggers them to question what they are being told. As detailed above, the present results revealed participants were in the truth-default state at the beginning of conversations, and certain aspects of vishing conversations triggered them into questioning the visher's honesty. Accordingly, our results

suggest TDT may be a useful tool for understanding how targeted individuals perceive vishers during vishing attacks.

Potential Applications of the Present Results

The present results have several important implications for the design of tools, such as digital assistants, that can advise targeted individuals during suspected vishing attacks. Specifically, the present results suggest such systems should reflect that people 1) begin conversations in the truth-default state, 2) are not concerned about seemingly innocuous requests, and 3) do not understand how seemingly innocuous requests can be used for nefarious purposes. One possibility would be to activate the system when the caller makes a request. Simply warning a user of potential deception could result in users leaving the truth-default state (Levine, 2014b), and additionally the system could convey information known to trigger suspicion and increase lie detection. For example, the system could convey i) a potential nefarious motive for the request (Bond Jr et al., 2013; Levine et al., 2010), and ii) questions to ask the caller which may reveal instances in which the caller says something that is inconsistent with earlier statements or with known reality (Levine, Blair, & Clare, 2014). Such features might shift users away from their truth-default state, increase their level of concern about seemingly innocuous requests, and help them understand how requests can be used for nefarious purposes. In doing so, the system might increase vishing attack detection.

The present results also have several important implications for penetration testing, i.e., authorized hacking to identify security vulnerabilities (Hadnagy, 2010). Specifically, the present results suggest penetration testers should 1) understand that people exhibit a truth-bias, so penetration testers should 2) aim to act in such a way so as to maintain that truth-bias. Toward that end, when possible, penetration testers should 3) aim to create contexts for their attacks that involve fairly typical situations, 4) request information that most people have not been warned against sharing, and 5) provide plausible explanations

for why requests for such information are reasonable given the circumstances. In doing so, penetration testers should increase the likelihood that their vishing attacks will be successful.

Future Research Directions

As the first study to examine perceptions of visher honesty during vishing attacks, the results of the present experiment lay the groundwork for future research in this important area. The present results highlight several aspects of vishing conversations that may cause honesty ratings to decrease, and other aspects of vishing conversations that may cause honesty ratings to increase. Future research should investigate whether those aspects of vishing conversations truly drive those effects. The present results also shed light on perceptions of visher honesty during vishing attacks, but do not speak to how those perceptions translate into target behavior. Future research should investigate the relationship between perceptions of visher honesty and whether targeted individuals comply with vishers' requests. Phishing research revealed that individual differences can predict phishing susceptibility (Lawson, Pearson, Crowson, & Mayhorn, 2020). Future research should investigate whether the same is true for vishing susceptibility.

Key Points

- People begin vishing conversations in the truth-default state.
- Certain aspects of vishing conversations serve as triggers.
- Other aspects of vishing conversations do not serve as triggers.
- In certain situations, people's perceptions of visher honesty improve.
- Truth-Default Theory may be a useful tool for understanding how targeted individuals behave during vishing attacks.

References

Baryshevtsev, M. V., et al. (2020). *Sharing is not caring: news features predict false news detection and diffusion* (Unpublished doctoral dissertation).

Bond Jr, C. F., Howard, A. R., Hutchison, J. L., & Masip, J. (2013). Overlooking the obvious: Incentives to lie. *Basic and Applied Social Psychology*, *35*(2), 212–221.

Bullée, J.-W. H., Montoya, L., Pieters, W., Junger, M., & Hartel, P. (2018). On the anatomy of social engineering attacks—A literature-based dissection of successful attacks. *Journal of Investigative Psychology and Offender Profiling*, *15*(1), 20–45.

Butavicius, M., Parsons, K., Pattinson, M., & McCormac, A. (2016). Breaching the human firewall: Social engineering in phishing and spear-phishing emails. *arXiv preprint arXiv:1606.00887*.

Carifio, J., & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about likert scales and likert response formats and their antidotes. *Journal of social sciences*, *3*(3), 106–116.

Clare, D. D., & Levine, T. R. (2019). Documenting the truth-default: The low frequency of spontaneous unprompted veracity assessments in deception detection. *Human Communication Research*, *45*(3), 286–308.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, *70*(5), 979.

Downs, J. S., Holbrook, M. B., & Cranor, L. F. (2006). Decision strategies and susceptibility to phishing. In *Proceedings of the second symposium on usable privacy and security* (pp. 79–90).

FBI, & IC3. (2019). 2018 Internet Crime Report. Retrieved from `https://pdf.ic3.gov/2018_IC3Report.pdf`

Feeley, T., & DeTurck, M. (1997). Perceptions of communications as seen by the actor and as seen by the observer: The case of lie detection. *International Communication Association Annual Conference*.

FTC. (2019). Consumer sentinel network data book for january-december 2018. Retrieved from https://www.ftc.gov/reports/consumer-sentinel-network-data-book-2018

Furnell, S. (2007). Phishing: can we spot the signs? *Computer Fraud & Security*, *2007*(3), 10–15.

Get Safe Online. (2015a, Aug 26). *Reconstruction of actual vishing call to a small business - Alfie Johnson.* YouTube. Retrieved from https://www.youtube.com/watch?v=IyzP1PIch8Y&list=PLT2O8mGYzrBOke-krvY2h9TOMPjpuxiWH&index=3

Get Safe Online. (2015b, Aug 26). *Reconstruction of actual vishing call to a small business - Archie Hicks.* YouTube. Retrieved from https://www.youtube.com/watch?v=lwc9iU2MidQ&list=PLT2O8mGYzrBOke-krvY2h9TOMPjpuxiWH

Get Safe Online. (2015c, Aug 26). *Reconstruction of actual vishing call to a small business - Harwood Estates.* YouTube. Retrieved from https://www.youtube.com/watch?v=MIgSM-7miVM&list=PLT2O8mGYzrBOke-krvY2h9TOMPjpuxiWH&index=2

Granhag, P. A., & Strömwall, L. A. (2001). Deception detection: Interrogators' and observers' decoding of consecutive statements. *The Journal of Psychology*, *135*(6), 603–620.

Greene, K. K., Steves, M. P., Theofanos, M. F., & Kostick, J. (2018). User context: an explanatory variable in phishing susceptibility. In *in proc. 2018 workshop usable security.*

Griffin, S. E., & Rackley, C. C. (2008). Vishing. *Proceedings of the 5th Annual Conference on Information Security Curriculum Development*, 33–35.

Hadnagy, C. (2010). *Social engineering: The art of human hacking.* John Wiley & Sons.

Havlicek, L. L., & Peterson, N. L. (1974). Robustness of the t test: A guide for researchers on effect of violations of assumptions. *Psychological Reports*, *34*(3_suppl), 1095–1114.

Heartfield, R., Loukas, G., & Gan, D. (2016). You are probably not the weakest link:

567 Towards practical prediction of susceptibility to semantic social engineering attacks.
568 *IEEE Access*, *4*, 6910–6928.
569 Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A
570 practical primer for t-tests and anovas. *Frontiers in Psychology*, *4*, 863.
571 Lawson, P., Pearson, C. J., Crowson, A., & Mayhorn, C. B. (2020). Email phishing and
572 signal detection: How persuasion principles and personality influence response
573 patterns and accuracy. *Applied Ergonomics*, *86*, 103084.
574 Levine, T. R. (2014a). Active deception detection. *Policy Insights from the Behavioral and*
575 *Brain Sciences*, *1*(1), 122–128.
576 Levine, T. R. (2014b). Truth-default theory (TDT): A theory of human deception and
577 deception detection. *Journal of Language and Social Psychology*, *33*(4), 378–392.
578 Levine, T. R. (2019). *Duped: Truth-default theory and the social science of lying and*
579 *deception*. University Alabama Press.
580 Levine, T. R., Blair, J. P., & Clare, D. D. (2014). Diagnostic utility: Experimental
581 demonstrations and replications of powerful question effects in high-stakes deception
582 detection. *Human Communication Research*, *40*(2), 262–289.
583 Levine, T. R., Kim, R. K., & Blair, J. P. (2010). (In)accuracy at detecting true and false
584 confessions and denials: An initial test of a projected motive model of veracity
585 judgments. *Human Communication Research*, *36*(1), 82–102.
586 Levine, T. R., Serota, K. B., Shulman, H., Clare, D. D., Park, H. S., Shaw, A. S., ... Lee,
587 J. H. (2011). Sender demeanor: Individual differences in sender believability have a
588 powerful impact on deception detection judgments. *Human Communication*
589 *Research*, *37*(3), 377–403.
590 MadeInSyr. (2013, Aug 5). *Live hack and social engineering at DEF_CON by Dave*
591 *Kennedy and Kevin Mitnick*. YouTube. Retrieved from
592 `https://youtu.be/DB6ywr9fngU?t=276`
593 Maggi, F. (2010). Are the con artists back? A preliminary analysis of modern phone

frauds. *2010 10th IEEE International Conference on Computer and Information Technology*, 824–831.

McCornack, S. A., Levine, T. R., Solowczuk, K. A., Torres, H. I., & Campbell, D. M. (1992). When the alteration of information is viewed as deception: An empirical test of information manipulation theory. *Communications Monographs*, *59*(1), 17–29.

Mitnick, K. D., & Simon, W. L. (2011). *The art of deception: Controlling the human element of security*. John Wiley & Sons.

Ollmann, G. (2007). The vishing guide. *IBM, Tech. Rep.*

Pandit, S., Perdisci, R., Ahamad, M., & Gupta, P. (2018). Towards measuring the effectiveness of telephony blacklists. In *Ndss*.

Park, H. S., & Levine, T. (2001). A probability model of accuracy in deception detection experiments. *Communication Monographs*, *68*(2), 201–210.

Parsons, K., McCormac, A., Pattinson, M., Butavicius, M., & Jerram, C. (2015). The design of phishing studies: Challenges for researchers. *Computers & Security*, *52*, 194–206.

Phelps, J., Nowak, G., & Ferrell, E. (2000). Privacy concerns and consumer willingness to provide personal information. *Journal of Public Policy & Marketing*, *19*(1), 27–41.

Power, R., & Forte, D. (2006). Social engineering: Attacks have evolved, but countermeasures have not. *Computer Fraud & Security*, *2006*(10), 17–20.

Proofpoint. (2019). *State of the phish: 2019 report* (Tech. Rep.). Retrieved from `https://info.wombatsecurity.com/hubfs/Wombat_Proofpoint_2019%20State%20of%20the%20Phish%20Report_Final.pdf`

Sawyer, B. D., & Hancock, P. A. (2018). Hacking the human: the prevalence paradox in cybersecurity. *Human factors*, *60*(5), 597–609.

Schindler, S., & Reinhard, M.-A. (2015). Increasing skepticism toward potential liars: Effects of existential threat on veracity judgments and the moderating role of honesty norm activation. *Frontiers in Psychology*, *6*, 1312.

Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The prevalence of lying in America: Three studies of self-reported lies. *Human Communication Research*, *36*(1), 2–25.

Street, C. N. (2015). Alied: Humans as adaptive lie detectors. *Journal of Applied Research in Memory and Cognition*, *4*(4), 335–343.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5). Pearson Boston, MA.

Tu, H., Doupé, A., Zhao, Z., & Ahn, G.-J. (2019). Users really do answer telephone scams. In *28th {USENIX} security symposium ({USENIX} security 19)* (pp. 1327–1340).

Vrij, A., Granhag, P. A., & Mann, S. (2010). Good liars. *The Journal of Psychiatry & Law*, *38*(1-2), 77–98.

Waugh, J. D., Glumm, M. M., Kilduff, P. W., Tauson, R. A., Smyth, C. C., & Pillalamarri, R. S. (2000). Cognitive workload while driving and talking on a cellular phone or to a passenger. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*(33), 6–276.

Biographies

Miriam E. Armstrong is a doctoral candidate in Texas Tech University's Human Factors program.

Keith S. Jones is an Associate Professor of Psychology at Texas Tech University. He received his Ph.D. in Experimental Psychology with an emphasis on Human Factors Psychology from the University of Cincinnati in 2000.

Akbar Siami Namin is an Associate Professor of Computer Science at Texas Tech University. He received his Ph.D. in Computer Science with an emphasis on Software Engineering from the University of Western Ontario in 2008.