Information & Computer Security

267

The effects of persuasion principles on perceived honesty during shoulder surfing attacks

Keith S. Jones, McKenna K. Tornblad, Miriam E. Armstrong and Jinwoo Choi
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 5 July 2023
Revised 16 May 2024
3 July 2024
Accepted 3 July 2024

Abstract

Purpose – This study aimed to investigate how honest participants perceived an attacker to be during shoulder surfing scenarios that varied in terms of which Principle of Persuasion in Social Engineering (PPSE) was used, whether perceived honesty changed as scenarios progressed, and whether any changes were greater in some scenarios than others.

Design/methodology/approach – Participants read one of six shoulder surfing scenarios. Five depicted an attacker using one of the PPSEs. The other depicted an attacker using as few PPSEs as possible, which served as a control condition. Participants then rated perceived attacker honesty.

Findings – The results revealed honesty ratings in each condition were equal during the beginning of the conversation, participants in each condition perceived the attacker to be honest during the beginning of the conversation, perceived attacker honesty declined when the attacker requested the target perform an action that would afford shoulder surfing, perceived attacker honesty declined more when the Distraction and Social Proof PPSEs were used, participants perceived the attacker to be dishonest when making such requests using the Distraction and Social Proof PPSEs and perceived attacker honesty did not change when the attacker used the target's computer.

Originality/value – To the best of the authors' knowledge, this experiment is the first to investigate how persuasion tactics affect perceptions of attackers during shoulder surfing attacks. These results have important implications for shoulder surfing prevention training programs and penetration tests.

Keywords Cybersecurity, Social engineering, Shoulder surfing, Persuasion, Authority, Commitment, Reciprocation and consistency, Distraction, Liking, similarity and deception, Social proof, Truth-default theory, Perceived honesty

Paper type Research paper

Introduction

Social engineering is the act of manipulating a person to take an action such as divulging sensitive information, e.g. passwords (Hadnagy, 2011). Social engineering is a common and costly cybersecurity threat. For example, a survey revealed that 85% of organizations

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.

2056-4961
DOI 10.1108/ICS-07-2023-0118

experienced a social engineering attack, with an average annual cost of over $1.4m per organization (Bissell and Ponemon, 2019). Such cost stems from business disruption, information loss and revenue loss, which concern losses in productivity and business processes, of sensitive or confidential information and of income from business opportunities due to a damaged reputation, respectively.

One form of social engineering is shoulder surfing, in which an attacker looks over their target's shoulder to steal sensitive information. The following provides a hypothetical shoulder surfing scenario, which is based on an attack described in Mitnick and Simon (2003) and similar to an attack described in Wang *et al.* (2021). Please see Bullée *et al.* (2015), Sheikh (2020) and Wang *et al.* (2021) for examples of attackers using similar approaches to breach organizations' physical security measures.

In our example scenario, an attacker enters an office building dressed in formal attire and accompanied by assistants. The attacker approaches the target and describes a fictitious scenario that will set the stage for the rest of the interaction, which is referred to as pretexting (Hadnagy, 2011). Specifically, the attacker introduces himself to the target in a friendly manner, notes that he is the company's Senior Accounts Manager and states that he is there for an important meeting. The attacker then tells the target that he needs to use the target's computer to send an important contract to the company's Chief Executive Officer (CEO). The target is hesitant to allow the attacker to do so. Accordingly, the attacker politely assures the target that he is who he says he is and shows the target a fake identification card and fictitious communications between himself and the CEO. The target then encourages the attacker to come around behind the target's desk to use the target's computer. From that vantage point, the attacker looks over the target's shoulder as the target enters their username and password. The attacker memorizes the target's credentials and later uses those credentials to break into the company's computer system.

Such attacks leverage human tendencies, such as our tendency to trust what others tell us (Longtchi *et al.*, 2024; Steinmetz *et al.*, 2021; Jampen *et al.*, 2020; Aldawood and Skinner, 2019) and to find people who act in certain ways to be very persuasive (Yasin *et al.*, 2021; Ferreira and Teles, 2019; Ferreira and Jakobsson, 2016; Ferreira and Lenzini, 2015; Ferreira *et al.*, 2015). Those tendencies are so engrained that social engineering experts are susceptible to such attacks (Hadnagy, 2011), as are individuals who were trained to guard against them (Adil *et al.*, 2020; Bullée and Junger, 2020). Accordingly, psychological research regarding deception detection and persuasion can help us understand why such shoulder surfing attacks are successful.

Deception detection

Several deception detection theories exist (for a review, see Masip, 2017). Truth-Default Theory (TDT) is one of the most well-supported theories (Levine, 2014a; 2014b; 2017; Serota *et al.*, 2021) and has been useful for understanding social engineering (Armstrong *et al.*, 2023).

According to TDT, people assume conversation partners are honest unless something "triggers" them to think otherwise (Levine, 2014b). Conversation partners are usually honest (Serota *et al.*, 2021) and lies are typically innocuous (Serota *et al.*, 2021), so it is generally adaptive to assume communication partners are honest (Levine, 2020). Potential triggers include a third-party's warning about potential deception, as well as conversation partners having an obvious motivation for deception, saying something that contradicts either something they said earlier or something the person knows to be true, or lacking an honest demeanor (Levine, 2014b). An honest demeanor includes confidence and composure, a pleasant, friendly, engaged and involved interaction style, and giving plausible explanations. A

dishonest demeanor includes avoiding eye contact; excessive fidgeting; appearing tense, nervous and anxious; an inconsistent interaction style; and speaking in a hesitant, uncertain and slow manner (Levine *et al.*, 2011). Once triggered, people search for evidence to confirm their suspicions. For example, they may look to previously acquired knowledge or experiences to evaluate the veracity of the other person's statements (Levine, 2014a). If they find sufficient evidence, then they will think their communication partner is being dishonest. Otherwise, they will revert to assuming their communication partner is being honest (Levine, 2014b).

Persuasion

There are numerous persuasion theories (for a review, see Cameron, 2009). They reference various persuasion principles, some of which are relevant to social engineering (Cialdini, 2007; Stajano and Wilson, 2011; Gragg, 2003). Ferreira *et al.* (2015) compiled a list of relevant persuasion principles and merged conceptually similar principles together to create a list of five *Principles of Persuasion in Social Engineering* (PPSE). Their Authority PPSE states that people tend to comply when they think the person making the request is an authority figure. Their Commitment, Reciprocation and Consistency PPSE states that people tend to comply when doing so jibes with a decision to which they have publicly committed, involves repaying a favor, or is consistent with their past behavior. Their Distraction PPSE states that people tend to comply when they focus on certain emotionally evocative aspects of an interaction, such as the urgent need to capitalize on a rare opportunity. That focus, and the associated emotional reaction, reduce their ability to think critically. Their Liking, Similarity and Deception PPSE states that people tend to comply when they perceive the person making the request to be likable, similar to themselves, familiar, or attractive. Their Social Proof PPSE states that people tend to comply when they think others have performed the requested behavior, others share any risks associated with the requested behavior or both.

Shoulder surfing through a psychological lens

The following revisits the shoulder surfing example discussed earlier, but with an eye toward how what is known about TDT and the PPSEs relates to the discourse. Our aim is to convey how TDT and the PPSEs can help us understand why shoulder surfing attacks are successful.

At the beginning of the example, the attacker entered the building dressed in business attire and accompanied by assistants. The attacker's clothing and entourage should convey a sense of authority. Given what is known about the PPSEs, such use of the Authority PPSE should encourage the target to comply with the attacker's request.

The attacker then approached the target, introduced himself to the target in a friendly manner, noted that he is the company's Senior Accounts Manager and stated that he was there for an important meeting. Given what is known about TDT, the target should think the attacker is being honest, unless something about the pretexting triggers the target to become suspicious. A potential trigger could be the target knowing the company's Senior Accounts Manager, i.e. the attacker saying something that contradicts something the target knows to be true. The attacker speaking in a friendly manner should indicate an honest demeanor and increase the likelihood that the target will think the attacker is honest. Furthermore, the content of the conversation should reinforce the target's sense that the attacker is an authority figure. Given what is known about the PPSEs, such use of the Authority PPSE should make it easier for the attacker to persuade the target to provide the attacker an opportunity to shoulder surf.

The attacker then told the target that he needs to use the target's computer to send an important contract to the company's CEO. Given what is known about TDT, the target should continue to think the attacker is being honest, unless something about the request or how it was made triggers suspicion. In addition, the attacker stating that the contract is

important and is being sent to the company's CEO should reinforce the air of authority surrounding the attacker and could cause the target to worry about potential consequences if the target does not allow the attacker to use their computer. Given what is known about the PPSEs, such as use of the Authority and Distraction PPSEs should increase the likelihood the target will comply with the attacker's request.

Nevertheless, the target was initially hesitant to comply with the attacker's request. In response, the attacker politely assured the target that he is who he says he is and provided evidence that supported his pretext. Given what is known about TDT, the attacker's politeness and their ability to substantiate the plausibility of their pretext should assuage any concerns the target might have had about the attacker.

The attacker then went behind the target's desk, looked over the target's shoulder as they entered their username and password and memorized the target's credentials. In doing so, the attacker gained information they needed to break into the company's computer system, and the target was none the wiser.

Use of certain PPSEs may trigger targets more so than use of other PPSEs

The preceding content makes it clear that the combination of TDT and PPSEs can help us better understand why shoulder surfing attacks are successful. In addition, that combination suggests that the use of certain PPSEs might trigger targets more so than the use of other PPSEs.

For example, consider the use of the Authority PPSE. Based on TDT, the use of this PPSE may trigger targets less than the use of other PPSEs. Use of the Authority PPSE will include tactics used to convince a target that the attacker is an authority figure. This may involve behaviors such as a confident demeanor, engaged interaction style and impression of knowledge when interacting with a target, which would suggest an honest demeanor (Levine *et al.*, 2011). Thus, when an attacker uses the Authority PPSE, the target may not be triggered as they attribute the attacker's behavior to their presumed authority role.

Conversely, consider the use of the Distraction PPSE. Based on TDT, the use of this PPSE may trigger targets more than the use of other PPSEs. Use of the Distraction PPSE will include tactics used to distract a target from other aspects of an interaction that may have large consequences, e.g. risk of losing confidential information. This may involve behaviors such as acting tense or nervous to convey a sense of urgency, which would suggest a dishonest demeanor (Levine *et al.*, 2011). Thus, when an attacker uses the Distraction PPSE, the target may be triggered as they attribute the attacker's behaviors to them being dishonest.

If true, such effects would have important implications for cybersecurity risk prevention. For example, shoulder surfing training could educate people about the fact that they are particularly susceptible when attackers leverage certain PPSEs, teach them how to recognize such situations and how to counteract the leveraged PPSE (Schaab *et al.*, 2017). In addition, penetration testing, which is a simulated attack on a system to analyze its security (Denis *et al.*, 2016), could leverage PPSEs that are not strong triggers during shoulder surfing attacks to increase the likelihood of the tester breaching the system's defenses.

The current experiment

Participants read a description of a shoulder surfing scenario. It either described a shoulder surfing scenario in which the attacker used one of the five PPSEs and did not use the other PPSEs or described the same shoulder surfing scenario but the attacker used as few PPSE-related behaviors as possible. The latter served as a control condition.

Each description was divided into five segments. The first segment, hereafter referred to as the Setting segment, describes the general setting of the story, with no mention of the attacker. The second segment, hereafter referred to as the Beginning segment, introduces the attacker and

describes the beginning of the attacker's conversation with the target. The third segment, hereafter referred to as the Request segment, describes the attacker requesting to use the target's computer. The fourth segment, hereafter referred to as the Compliance segment, describes the target complying with the attacker's request. The fifth segment, hereafter referred to as the Conclusion segment, describes the end of the attacker's interaction with the target. Participants rated attacker honesty after reading each segment, except for the Setting segment.

The present experiment addressed six research questions:

RQ1. Were honesty ratings in each condition equal during the beginning of the conversation (Beginning segment)?

RQ2. Did participants in each condition perceive the attacker to be honest, neutral or dishonest during the beginning of the conversation (Beginning segment)?

RQ3. Did honesty ratings decline when the attacker made the request (between the Beginning and Request segments)?

RQ4. Did honesty ratings decline in certain conditions more than others when the attacker made the request (between the Beginning and Request segments)?

RQ5. Did participants in each condition perceive the attacker to be honest, neutral, or dishonest during the request (Request segment)?

RQ6. Did honesty ratings change when the attacker used the target's computer (between the Request and Compliance segments)?

Based on TDT, we predicted that honesty ratings in each condition would be equal at the beginning of the conversation (*RQ1)* and that participants would perceive the attacker to be honest at the beginning of their conversation with the target (*RQ2)*. Finally, we predicted that a request to use the target's computer would be a trigger (*RQ3)*. We did not make specific predictions for the remaining research questions because we thought the possibilities described in the "Use of Certain PPSEs May Trigger Targets More So Than Use of Other PPSEs" section, although plausible, were too speculative.

The present experiment was the first to investigate whether using certain PPSEs during shoulder surfing attacks would trigger targets more so than use of other PPSEs. The results of the present experiment have important implications for the development of shoulder surfing prevention training programs and for conducting penetration tests.

Method

Participants

In total, 108 members of the campus community participated in the study. They were recruited through a university announcement system and fliers distributed throughout campus and paid $12 for their participation. Eighteen participants had missing data or responded carelessly and were removed from the data set (see the Data Cleaning section for details). The resultant sample contained 90 participants (33 male, 56 female, 1 other; M_{age} = 23.76, SD_{age} = 6.97).

Experimental design

The present study used a one within- one between-subjects design. The within-subjects variable was Description Segment, which included five levels: the (1) Setting, (2) Beginning, (3) Request, (4) Compliance, and (5) Conclusion segments. The between-subjects variable was PPSE Condition, which included six conditions: (1) Control, (2) Authority,

(3) Commitment, Reciprocation and Consistency, (4) Distraction, (5) Liking, Similarity and Deception, and (6) Social Proof.

The dependent variable was perceived attacker honesty, which was the mean rating across four semantic difference items: (1) misleading/not misleading, (2) deceitful/truthful, (3) dishonest/honest, and (4) deceptive/not deceptive (α = 0.95). Those items were used in previous research concerning manipulation (McCornack *et al.*, 1992).

Materials

Hardware. The experimenter's workstation was a 2020 MacBook Air (Apple M1 Processor, 8GB RAM, macOS 12 Monterey). The participants' workstations were their personal computers, so their specifications varied.

Software. Two software packages were used: Zoom and Qualtrics. The former allowed the experimenter and participant to communicate; the latter afforded study administration.

Descriptions. Six descriptions were used. Each described a shoulder surfing scenario from the target's point of view and was divided into five segments. The Setting segment described the general background and setting of the story. The Beginning segment described the attacker, David Johnson, introducing himself and beginning to converse with the target. The Request segment described the attacker asking to use the target's computer. The Compliance segment described the target complying with the attacker's request and the attacker's subsequent use of the target's computer. The Conclusion segment described the attacker thanking the target for their compliance.

The Control description included as few PPSE-related elements as possible. Each other description discussed the attacker using one of the PPSEs. Elements related to the target PPSE occurred in the Beginning, Request, Compliance and Conclusion segments in narrative-appropriate locations.

Attacker questionnaire. This questionnaire included the prompt, "How would you describe David Johnson in the paragraph above?", which was followed by a list containing 14 semantic difference items: (1) misleading/not misleading, (2) deceitful/truthful, (3) dishonest/honest, (4) deceptive/not deceptive, (5) trained/untrained, (6) experienced/inexperienced, (7) skilled/unskilled, (8) qualified/unqualified, (9) informed/uninformed, (10) aggressive/meek, (11) emphatic/hesitant, (12) bold/timid, (13) active/passive, and (14) energetic/tired. The first four, next five and last five items comprised the perceived honesty scale (McCornack *et al.*, 1992), a qualification scale (Berlo *et al.*, 1969) and a dynamism scale (Berlo *et al.*, 1969), respectively. The latter two scales served as filler. Each item included a six-point scale (e.g. dishonest = 1; honest = 6). The order of the items was randomized for each description segment that was rated, i.e. the Beginning, Request, Compliance and Conclusion segments.

Procedure

Description development and validation. The core description narrative was inspired by an exchange between an attacker and a receptionist that was described in Mitnick and Simon (2003). To create the Control description, that narrative was modified to describe a shoulder surfing attack, remove as many PPSE-related elements as possible and ensure the narrative remained coherent after those elements were changed or removed. To create each PPSE description, the Control description's narrative was modified to describe the attacker using the target PPSE.

To refine and validate the descriptions, a series of manipulation check experiments was conducted. In each experiment, manipulation check participants read one of the descriptions and then rated the extent to which PPSE-related factors compelled the receptionist to comply with the attacker's request. Items were based on Ferreira and Lenzini's (2015) PPSE

descriptions. Each item was rated on a six-point scale (e.g. Not at all = 1; To a great extent = 6). Ratings for each PPSE description were compared against ratings for the Control description. When either ratings for the target PPSE were not significantly different than ratings for that PPSE in the Control description or ratings for a nontarget PPSE were significantly different than ratings for those PPSEs in the Control description, narrative elements in the PPSE description, the Control description or both were modified, and another manipulation check experiment was conducted. That process was repeated until, for each PPSE description, ratings for the target PPSE were significantly different than ratings for that PPSE in the Control description and ratings for nontarget PPSEs were not significantly different than ratings for those PPSEs in the Control description.

The sole exception was the Commitment, Reciprocation and Consistency comparison between the Control and Authority descriptions. Commitment, Reciprocation and Consistency ratings were significantly greater in the Control description than the Authority description. Despite our best efforts, we were unable to eliminate that difference and concluded that something integral to how we implemented the Authority PPSE in the Authority description must have lowered Commitment, Reciprocation, and Consistency ratings below baseline. Thankfully, that difference was not an issue in the present experiment because we did not observe any significant differences in perceived honesty between the Control and Authority PPSE conditions.

Data collection. Each participant completed the experiment individually. Testing sessions lasted approximately 20 min and were conducted via Zoom due to the COVID-19 pandemic.

The experimenter emailed the participant a link for a Zoom meeting. The participant joined that meeting. The experimenter then sent the participant a link to a Qualtrics study via Zoom. The participant followed that link to start the experiment.

The participant read and agreed to an informed consent statement and then read instructions. The participant was then assigned to one of the six description conditions at random with the constraint that a participant would be assigned to each of the six conditions before another participant would be assigned to a previously assigned condition. The participant then read the first segment of their assigned description. The participant then read the second description segment and completed the Attacker Questionnaire. The participant repeated that process for the remaining description segments. The participant was then debriefed.

Data cleaning

The data were examined for missing responses and careless responding. Eight participants had missing data for at least one description segment. It was not possible to impute values for those missing responses, so those participants were removed from the sample. Ten participants exhibited careless responding. Specifically, those participants' responses to three of the four perceived attacker honesty items were on one end of the six-point scale whereas the remaining response was on the opposite end of the scale. For example, a participant rated misleading/not misleading, deceitful/truthful and dishonest/honest as a "1" on the six-point scale and rated deceptive/not deceptive as a "6" on that scale. In other words, that participant rated David Johnson as being misleading, deceitful and dishonest, as well as not deceptive. Such wide variation between responses suggested careless responding, so those participants were removed from the sample. We speculate that careless responders volunteered solely to get paid and thus did not fully engage with the study.

Results and discussion

Analytic approach

The following subsections detail tests performed to answer our 6 research questions. The order of the subsections follows the flow of the conversations described in the descriptions.

For each test, we performed independent samples *t*-tests, paired samples *t*-tests or one sample *t*-tests, depending on the research question. We used parametric tests because our honesty ratings were composite scores derived from response items with discrete values and thus likely exhibited interval scale characteristics (Carifio and Perla, 2007), and even if they did not, *t*-tests are robust when the interval-level data assumption is violated (Carifio and Perla, 2007; Havlicek and Peterson, 1974). Reported effect sizes are Cohen's d_s for independent samples *t*-tests, d_{rm} for paired samples *t*-tests and d_z for one sample *t*-tests (Lakens, 2013).

One sample *t*-tests evaluated whether participants perceived the attacker to be honest, neutral or dishonest. To do so, they compared honesty ratings against the honesty scale's neutral point, that is, 3.5. We considered participants to perceive the attacker to be honest, neutral or dishonest when their honesty ratings were significantly greater than 3.5, not significantly different than 3.5, or significantly lesser than 3.5, respectively.

We considered the set of *t*-tests associated with each research question to be a family. We applied the Bonferroni correction so that family-wise $\alpha = 0.05$. See Table 1 for descriptive statistics for each condition in the Beginning, Request and Compliance segments, respectively.

Were honesty ratings in each condition equal during the beginning of the conversation (Beginning segment)?

We conducted 15 independent sample *t*-tests ($\alpha = 0.05/15 = 0.003$) to determine whether honesty ratings in each condition were equal during the beginning of the conversation. There was one *t*-test for each of the possible PPSE comparisons (e.g. Control vs Authority). None of the *t*-tests yielded significant differences, indicating that all conditions began with participants perceiving the attacker's honesty level during the beginning of the conversation in similar ways. See Table 2 for details.

We predicted perceived honesty for the Beginning segment would be consistent across conditions. The present results support that prediction.

Did participants in each condition perceive the attacker to be honest, neutral, or dishonest during the beginning of the conversation (Beginning segment)?

We conducted six one-sample *t*-tests ($\alpha = 0.05/6 = 0.008$) to determine whether participants in each condition perceived the attacker to be honest, neutral or dishonest during the

Table 1. Honesty ratings for each vignette segment

PPSE	Beginning segment Mean (SD)	Request segment Mean (SD)	Compliance segment Mean (SD)
Control	5.17 (0.72)	4.18 (1.35)	4.28 (1.63)
Authority	4.55 (0.95)	2.70 (1.22)	2.83 (1.03)
Commitment, reciprocation and consistency	4.60 (0.74)	3.27 (1.03)	3.08 (1.11)
Distraction	5.20 (1.11)	2.07 (0.96)	2.00 (1.13)
Liking, similarity and deception	4.50 (0.88)	3.07 (1.20)	2.93 (1.32)
Social proof	5.07 (0.92)	2.37 (1.08)	2.48 (1.23)

Source: Authors' own creation

Table 2. Comparisons of honesty ratings between conditions during the beginning of the conversation (Beginning segment)

Comparison	Test statistic	*p*-value	Cohen's d_s
Control – Authority	2.00	0.055	0.73
Control – Commitment, Reciprocation and Consistency	2.12	0.043	0.77
Control – Distraction	−0.10	0.923	−0.04
Control – Liking, Similarity and Deception	2.26	0.032	0.83
Control – Social Proof	0.33	0.743	0.12
Authority – Commitment, Reciprocation and Consistency	−0.16	0.874	−0.06
Authority – Distraction	−1.73	0.095	−0.63
Authority – Liking, Similarity and Deception	0.15	0.882	0.05
Authority – Social Proof	−1.51	0.141	−0.55
Commitment, Reciprocation and Consistency – Distraction	−1.74	0.092	−0.64
Commitment, Reciprocation and Consistency – Liking, Similarity and Deception	0.34	0.739	0.12
Commitment, Reciprocation and Consistency – Social Proof	−1.53	0.137	0.37
Distraction – Liking, Similarity and Deception	1.92	0.066	0.70
Distraction – Social Proof	0.36	0.722	0.13
Liking, Similarity and deception – Social Proof	−1.72	0.096	−0.63

Source: Authors' own creation

beginning of the conversation (Beginning segment). Each of those *t*-tests revealed that honesty ratings were significantly greater than the honesty scale's neutral point of 3.5, which indicates that participants in all six conditions perceived the attacker to be honest during the beginning of the conversation. See Table 3 for details.

We predicted participants would perceive the attacker to be honest at the beginning of their conversation with the target. The present results support that prediction and replicate those reported by Armstrong *et al.* (2023), which concerned vishing rather than shoulder surfing.

Participants likely perceived the attacker to be honest because there is a very strong tendency for people to enter conversations assuming that their partner is being honest (Levine, 2014b). Furthermore, the Beginning segments of each of the descriptions involve the attacker

Table 3. Comparisons of honesty ratings against the honesty scale's neutral point (3.5) during the beginning of the conversation (Beginning segment)

Condition	Test statistic	*p*-value	Cohen's d_z
Control	*8.92*	*<0.001*	*2.30*
Authority	*4.28*	*<0.001*	*1.10*
Commitment, Reciprocation and Consistency	*5.74*	*<0.001*	*1.48*
Distraction	*5.95*	*<0.001*	*1.54*
Liking, Similarity and Deception	*4.39*	*<0.001*	*1.13*
Social Proof	*6.61*	*<0.001*	*1.71*

Note: Rows with statistically significant differences are italiced
Source: Authors' own creation

exhibiting an honest demeanor. Specifically, the Beginning segment of each description stated the attacker politely introducing himself to the target and then engaging in conversation with the target. As such, the attacker exhibited a pleasant and engaged interaction style, which are known characteristics of an honest demeanor (Levine *et al.*, 2011).

Did honesty ratings decline when the attacker made the request (between the Beginning and Request segments)?

We conducted six paired sample *t*-tests ($\alpha = 0.05/6 = 0.008$) to determine whether honesty ratings declined when the attacker asked for access to the target's computer. Those tests compared honesty ratings from the Beginning and Request segments for each condition. All six paired sample *t*-tests revealed that honesty ratings for the Request segment were significantly less than those for the Beginning segment, which indicates that honesty ratings in each condition declined when the attacker made the request. See Table 4 for details.

These results likely reflect the sensitive nature of the attacker's request, i.e. asking for access to the target's computer. Requests for highly sensitive information served as triggers during phishing attacks (Downs *et al.*, 2006; Furnell, 2007) and vishing attacks (Armstrong *et al.*, 2023). As such, the present results replicate those findings. Further, most workers have been explicitly told that they should not let others use their computer. Therefore, the attacker's request may have served as a trigger because the attacker violated what participants were taught or because the request caused participants to consider the attacker's motivations for requesting to use the target's computer (Levine, 2014b).

Did honesty ratings decline more in certain conditions than others when the attacker made the request (between the Beginning and Request segments)?

We conducted 15 independent sample *t*-tests ($\alpha = 0.05 / 15 = 0.003$) to determine whether honesty ratings declined in certain conditions more so than others when the attacker asked for access to the target's computer (between the Beginning and Request segments). Those *t*-tests compared differences in perceived honesty between the Beginning and Request segments across conditions. The results revealed that perceived honesty in the Distraction condition declined more than in the Control condition; in the Commitment, Reciprocation and Consistency condition; and in the Liking, Similarity and Deception condition. Furthermore, perceived honesty in the Social Proof condition declined more than in the Control condition. These results suggest that using the Distraction PPSE when making one's request caused honesty ratings to decline more so than using certain other PPSEs.

Table 4. Comparisons of honesty ratings between the beginning of the conversation (Beginning segment) and when the attacker made the request (Request segment)

Condition	Test statistic	*p*-value	Cohen's d_{rm}
Control	*3.14*	*0.007*	*0.85*
Authority	*5.69*	*<0.001*	*1.68*
Commitment, Reciprocation and Consistency	*4.98*	*<0.001*	*1.46*
Distraction	*8.08*	*<0.001*	*3.02*
Liking, Similarity and Deception	*4.92*	*<0.001*	*1.34*
Social Proof	*7.82*	*<0.001*	*2.70*

Note: Rows with statistically significant differences are italiced
Source: Authors' own creation

Furthermore, using the Social Proof PPSE caused honesty ratings to decline more so than using as few PPSE-related elements as possible. See Table 5 for details. The present results replicate those reported by Parsons *et al.* (2019), which revealed that people are most suspicious of phishing emails that use the Distraction and Social Proof PPSEs.

Why did perceived honesty decline more in the Distraction and Social Proof conditions than in certain other conditions? One possibility is that participants may have perceived the attacker in the Distraction and Social Proof descriptions as having a dishonest demeanor (Levine, 2014b), which amplified the decline in honesty ratings that would have otherwise occurred because the attacker asked to use the target's computer.

In the Distraction description, the attacker exhibited several behaviors that could have caused participants to perceive the attacker as having a dishonest demeanor. First, the attacker displayed an inconsistent interaction style in the Distraction description as opposed to a relatively consistent interaction style in the Commitment, Reciprocation and Consistency, Liking, Similarity and Deception and Control descriptions. In the Distraction description, the attacker interacts with the target in a calm and polite manner during the Beginning segment and in a rushed, anxious and relatively impolite manner during the Request segment. In the Commitment; Reciprocation and Consistency; Liking, Similarity and Deception; and Control descriptions, the attacker interacts with the target in a calm and polite manner during the Beginning segment and again in the Request segment. Second, the attacker acted tense, nervous and anxious in the Distraction description and polite in the Commitment; Reciprocation and Consistency; Liking, Similarity and Deception; and Control descriptions. In the Distraction description, the attacker talked in a fast, hushed tone when they told the target they needed to use the target's computer. In the Commitment; Reciprocation and Consistency; Liking, Similarity and Deception; and Control descriptions, the attacker talked in a polite manner when they asked the target whether they could use their

Table 5. Comparisons of mean differences in honesty ratings from the beginning of the conversation (Beginning segment) and when the attacker made the request (Request segment) across conditions

Comparison	Test statistic	*p*-value	Cohen's d_s
Control – Authority	−1.92	0.065	−0.70
Control – Commitment, Reciprocation and Consistency	−0.85	0.403	−0.31
Control – Distraction	*−4.31*	*<0.001*	*−1.58*
Control – Liking, Similarity and Deception	−1.05	0.301	−0.38
Control – Social Proof	*−3.68*	*<0.001*	*−1.35*
Authority – Commitment, Reciprocation and Consistency	1.23	0.230	0.45
Authority – Distraction	−2.53	0.017	−0.93
Authority – Liking, Similarity and Deception	0.95	0.348	0.35
Authority – Social Proof	−1.79	0.084	−0.65
Commitment, Reciprocation and Consistency – Distraction	*−3.82*	*<0.001*	*−1.39*
Commitment, Reciprocation and Consistency – Liking, Similarity and Deception	−0.25	0.802	−0.09
Commitment, Reciprocation and Consistency – Social Proof	−3.13	0.004	−1.14
Distraction – Liking, Similarity and Deception	*3.50*	*0.002*	*1.28*
Distraction – Social Proof	0.83	0.411	0.30
Liking, Similarity and Deception – Social Proof	−2.80	0.009	−1.02

Note: Rows with statistically significant differences are italiced
Source: Authors' own creation

computer. Individually or collectively, these factors may have caused participants in the Distraction condition to perceive the attacker as dishonest (Levine, 2014b).

In the Social Proof description, the attacker also exhibited several behaviors that could have caused participants to perceive the attacker as having a dishonest demeanor. First, the attacker mentions that "other receptionists have assisted [him] with this problem before," which seems openly coercive. Second, the attacker follows up their request to use the target's computer by saying, "If someone makes a fuss about letting me use your computer, you can just blame it on me. I'm sure your other visitors would back you up too." Here, the attacker tacitly acknowledges that they should not be using the receptionist's computer, which suggests the attacker expects to arouse some suspicion. Third, the attacker's assurance that other visitors would back up the target seems atypical and implausible because the attacker made no mention of the visitors beforehand. The implausibility of that statement may have led participants to perceive the attacker as dishonest (Levine *et al.*, 2011). Individually or collectively, these factors may have caused participants in the Social Proof condition to perceive the attacker as dishonest (Levine, 2014b).

Did participants in each condition perceive the attacker to be honest, neutral, or dishonest during the request? (Request segment)?

We conducted six one-sample *t*-tests ($\alpha = 0.05/6 = 0.008$) to determine whether participants in each condition perceived the attacker to be honest, neutral or dishonest during the request (Request segment). Four of the six one-sample *t*-tests did not reveal significant differences from the honesty scale's neutral point of 3.5, indicating that participants in these conditions were in a neutral state when the attacker asked for access to the target's computer. In addition, two of the six one-sample *t*-tests revealed that honesty ratings were significantly less than the honesty scale's neutral point of 3.5, indicating that participants in these conditions perceived the attacker to be dishonest when the attacker made the request. See Table 6 for details.

The present results revealed that participants in the Distraction and Social Proof conditions perceived the attacker as dishonest. The immediately preceding section discussed several reasons why that might be the case.

Did honesty ratings change when the attacker used the target's computer (between the Request to Compliance segments)?

We conducted six paired sample *t*-tests ($\alpha = 0.05/6 = 0.008$) to determine whether honesty ratings changed when the attacker used the target's computer. Those tests compared honesty

Table 6. Comparison of honesty ratings against the honesty scale's neutral point (3.5) when the attacker made the request (Request segment)

Condition	Test statistic	*p*-Value	Cohen's d_z
Control	1.96	0.070	0.51
Authority	−2.54	0.023	−0.66
Commitment, Reciprocation and Consistency	−0.88	0.396	−0.23
Distraction	*−5.78*	*<0.001*	*−1.49*
Liking, Similarity and Deception	−1.40	0.184	−0.36
Social Proof	*−4.08*	*0.001*	*−1.05*

Note: Rows with statistically significant differences are italiced
Source: Authors' own creation

ratings from the Request segment to those in the Compliance segment for each condition. None of those *t*-tests revealed significant differences between those segments, which indicates that honesty ratings in each condition did not change when the attacker used the target's computer. See Table 7 for details.

The present results are different than those reported by Armstrong *et al.* (2023), which investigated perceived attacker honesty during vishing attacks. Specifically, Armstrong *et al.* revealed that honesty ratings recovered slightly after requests for sensitive information. In contrast, the present results suggest that honesty ratings did not change between the Request and Compliance segments.

Armstrong *et al.* (2023) attributed the observed recovery to the fact that the attacker explained the reason for their request after making it. Specifically, they noted that providing follow-up information may have made the request seem plausible, which in turn would have made the attacker seem more honest than they were perceived to be before. In the present experiment, the attacker did not explain the reason for their request during the Compliance segment, which might explain why the honesty ratings in the present experiment did not recover when the attacker used the target's computer.

Practical implications

The present results revealed the following insights about shoulder surfing attacks: (1) perceived attacker honesty is equal in the beginning of a shoulder surfing attack, regardless of the PPSE used by the attacker, (2) people perceive the attacker to be honest during the beginning of a shoulder surfing attack, (3) perceived attacker honesty declines when the attacker requests the target to perform an action that will afford shoulder surfing, (4) perceived attacker honesty declines more when the Distraction and Social Proof PPSEs are used, (5) people perceive the attacker to be dishonest when making such requests using the Distraction and Social Proof PPSEs, and (6) perceived attacker honesty does not change during the request.

Advice for training programs

The results have important implications for the development of specific PPSE-related shoulder surfing prevention training programs. Such training should take into consideration that:

- people perceive the attacker to be honest during the beginning of the shoulder surfing attack; and

Table 7. Comparisons of honesty ratings between when the attacker made the request (Request segment) and when the attacker used the target's computer (Compliance segment)

Condition	Test statistic	*p*-value	Cohen's d_{rm}
Control	−0.50	0.624	−0.06
Authority	−0.51	0.615	−0.12
Commitment, Reciprocation and Consistency	0.99	0.338	0.17
Distraction	0.44	0.670	0.06
Liking, Similarity and Deception	0.45	0.658	0.11
Social Proof	−0.47	0.648	−0.10

Source: Authors' own creation

- that perceived attacker honesty declines more when the Distraction and Social Proof PPSEs are used.

Therefore, such training should emphasize human tendencies to perceive others as honest during potential shoulder surfing scenarios. Also, training should note that use of the Commitment; Reciprocation and Consistency; and Liking, Similarity and Deception PPSEs resulted in lesser declines in perceived attacker honesty relative to use of the Distraction PPSE, which could increase one's susceptibility to attacks that use the former PPSEs. Furthermore, training should stress how the Commitment; Reciprocation and Consistency; and Liking, Similarity and Deception PPSEs are implemented, provide examples of how they are used in real-world scenarios, and, when possible, demonstrate trainees' vulnerability to such PPSEs. Such activities should counteract the effects of those PPSEs (Schaab *et al.*, 2017).

Advice for penetration testing

The results also have important implications for the specific techniques used in penetration testing. Such testing should take into consideration that:

- perceived attacker honesty declines when the attacker requests the target to perform an action that will afford shoulder surfing;
- perceived attacker honesty declines more when the Distraction and Social Proof PPSEs are used;
- people perceive the attacker to be dishonest when making such requests using the Distraction and Social Proof PPSEs; and
- perceived attacker honesty does not change during the request.

In doing so, penetration testers will use the PPSEs to which people are most vulnerable, i.e. using the Commitment; Reciprocation and Consistency; and Liking, Similarity and Deception PPSEs rather than the Distraction and Social Proof PPSEs. Using PPSEs that result in less steep declines in perceived honesty when the request is made may increase the likelihood that they can breach the system.

Future research

We offered possible reasons why perceived attacker honesty declines more when the Distraction and Social Proof PPSEs are used compared to when other PPSEs are used. For example, we noted that the attacker's inconsistent interaction style in the Distraction description may have caused participants in that condition to perceive the attacker as having a dishonest demeanor (Levine, 2014b), which may have amplified the decline in honesty ratings that would have otherwise occurred because the attacker asked to use the target's computer. Future research should investigate such possibilities. To do so, one could repeat the present experiment, but with the addition of an alternative version of the Distraction description, in which the attacker attempts to distract their target but does so while maintaining a consistent interaction style. If our original Distraction description leads to a more substantial decline in perceived attacker honesty than the alternative Distraction description, then that would implicate the attacker's inconsistent interaction style. Similar studies should be conducted to investigate the other possibilities that we offered.

Future research should also investigate whether perceived attacker honesty declines more when the Distraction and Social Proof PPSEs are used during *vishing* attacks compared to when other PPSEs are used. To do so, one could revise the descriptions used in the present

experiment to reflect vishing rather than shoulder surfing scenarios and repeat the experiment. If the present results replicate, then that would suggest that the present results apply to vishing as well.

References

Adil, M., Khan, R. and Nawaz Ul Ghani, M.A. (2020), "Preventive techniques of phishing attacks in networks", 2020 3rd International Conference on Advancements in Computational Sciences (ICACS), doi: 10.1109/icacs47775.2020.9055943.

Aldawood, H. and Skinner, G. (2019), "A taxonomy for social engineering attacks via personal devices", *International Journal of Computer Applications*, Vol. 178 No. 50, pp. 19-26.

Armstrong, M.E., Jones, K.S. and Namin, A.S. (2023), "How perceptions of caller honesty vary during vishing attacks that include highly sensitive or seemingly innocuous requests", *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 65 No. 2, pp. 275-287, doi: 10.1177/00187208211012818.

Berlo, D.K., Lemert, J.B. and Mertz, R.J. (1969), "Dimensions for evaluating the acceptability of message sources", *Public Opinion Quarterly*, Vol. 33 No. 4, pp. 563-576, doi: 10.1086/267745.

Bissell, K. and Ponemon, L. (2019), "Accenture/Ponemon Institute: the cost of cybercrime", *Network Security*, Vol. 2019 No. 3, pp. 1-23, doi: 10.1016/s1353-4858(19)30032-7.

Bullée, J.-W. and Junger, M. (2020), "Social engineering", *The Palgrave Handbook of International Cybercrime and Cyberdeviance*, pp. 849-875.

Bullée, J.-W.H., Montoya, L., Pieters, W., Junger, M. and Hartel, P.H. (2015), "The persuasion and security awareness experiment: reducing the success of social engineering attacks", *Journal of Experimental Criminology*, Vol. 11 No. 1, pp. 97-115.

Cameron, K.A. (2009), "A practitioner's guide to persuasion: an overview of 15 selected persuasion theories, models and frameworks", *Patient Education and Counseling*, Vol. 74 No. 3, pp. 309-317, doi: 10.1016/j.pec.2008.12.003.

Carifio, J. and 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*, Vol. 3 No. 3, pp. 106-116, doi: 10.3844/jssp.2007.106.116.

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

Denis, M., Zena, C. and Hayajneh, T. (2016), "Penetration testing: concepts, attack methods, and defense strategies", 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT), pp. 1-6, doi: 10.1109/lisat.2016.7494156.

Downs, J.S., Holbrook, M.B. and Cranor, L.F. (2006), "Decision strategies and susceptibility to phishing", *Proceedings of the Second Symposium on Usable Privacy and Security – SOUPS'06*, doi: 10.1145/1143120.1143131.

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

Ferreira, A. and Jakobsson, M. (2016), "Persuasion in scams", *Understanding Social Engineering Based Scams*, pp. 29-47.

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.

Ferreira, A., Coventry, L. and Lenzini, G. (2015), "Principles of persuasion in social engineering and their use in phishing", *Lecture Notes in Computer Science*, pp. 36-47.

Furnell, S. (2007), "Phishing: can we spot the signs?", *Computer Fraud and Security*, Vol. 2007 No. 3, pp. 10-15, doi: 10.1016/s1361-3723(07)70035-0.

Gragg, D. (2003), *A Multi-Level Defense Against Social Engineering*, *SANS Institute*.

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

Havlicek, L.L. and Peterson, N.L. (1974), "Robustness of the *t* test: a guide for researchers on effect of violations of assumptions", *Psychological Reports*, Vol. 34 No. 3_suppl, pp. 1095-1114, doi: 10.2466/pr0.1974.34.3c.1095.

Jampen, D., Gür, G., Sutter, T. and Tellenbach, B. (2020), "Don't click: towards an effective anti-phishing training. A comparative literature review", *Human-Centric Computing and Information Sciences*, Vol. 10 No. 1, doi: 10.1186/s13673-020-00237-7.

Lakens, D. (2013), "Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for T-tests and ANOVAS", *Frontiers in Psychology*, Vol. 4, doi: 10.3389/fpsyg.2013.00863.

Levine, T.R. (2014a), "Active deception detection", *Policy Insights from the Behavioral and Brain Sciences*, Vol. 1 No. 1, pp. 122-128, doi: 10.1177/2372732214548863. volNo.

Levine, T.R. (2014b), "Truth-default theory (TDT)", *Journal of Language and Social Psychology*, Vol. 33 No. 4, pp. 378-392, doi: 10.1177/0261927x14535916.

Levine, T.R. (2017), "Mysteries and myths in human deception and deception detection: insights from truth-default theory", *Ewha Journal of Social Sciences*, Vol. 33 No. 2, pp. 5-28, doi: 10.16935/ejss.2017.33.2.001.

Levine, T.R. (2020), *Duped: truth-Default Theory and the Social Science of Lying and Deception*, University of AL Press, Tuscaloosa, AL.

Levine, T.R., Serota, K.B., Shulman, H., Clare, D.D., Park, H.S., Shaw, A.S., … Lee, J.H. (2011), "Sender demeanor: individual differences in sender believability have a powerful impact on deception detection judgments", *Human Communication Research*, Vol. 37 No. 3, pp. 377-403, doi: 10.1111/j.1468-2958.2011.01407.x.

Longtchi, T.T., Rodriguez, R.M., Al-Shawaf, L., Atyabi, A. and Xu, S. (2024), "Internet-based social engineering psychology, attacks, and defenses: a survey", *Proceedings of the IEEE*, pp. 1-37.

McCornack, S.A., Levine, T.R., Solowczuk, K.A., Torres, H.I. and Campbell, D.M. (1992), "When the alteration of information is viewed as deception: an empirical test of information manipulation theory", *Communication Monographs*, Vol. 59 No. 1, pp. 17-29, doi: 10.1080/03637759209376246.

Masip, J. (2017), "Deception detection: state of the art and future prospects", *Psicothema*, Vol. 29 No. 2, pp. 149-159, doi: 10.7334/psicothema2017.34.

Mitnick, K.D. and Simon, W.L. (2003), *The Art of Deception: controlling the Human Element of Security*, Wiley, Indianapolis, IN.

Parsons, K., Butavicius, M., Delfabbro, P. and Lillie, M. (2019), "Predicting susceptibility to social influence in phishing emails", *International Journal of Human-Computer Studies*, Vol. 128, pp. 17-26.

Schaab, P., Beckers, K. and Pape, S. (2017), "Social engineering defence mechanisms and counteracting training strategies", *Information and Computer Security*, Vol. 25 No. 2, pp. 206-222.

Serota, K.B., Levine, T.R. and Docan-Morgan, T. (2021), "Unpacking variation in lie prevalence: prolific liars, bad lie days, or both?", *Communication Monographs*, Vol. 89 No. 3, pp. 307-331, doi: 10.1080/03637751.2021.1985153.

Sheikh, A. (2020), *Comptia Security+ Certification Study Guide: Network Security Essentials*, Apress L. P., Berkeley, CA.

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.

Steinmetz, K.F., Pimentel, A. and Goe, W.R. (2021), "Performing social engineering: a qualitative study of information security deceptions", *Computers in Human Behavior*, Vol. 124, doi: 10.1016/j.chb.2021.106930.

Wang, Z., Zhu, H. and Sun, L. (2021), "Social engineering in cybersecurity: effect mechanisms, human vulnerabilities and attack methods", *IEEE Access*, Vol. 9, pp. 11895-11910.

Yasin, A., Fatima, R., Liu, L., Wang, J., Ali, R. and Wei, Z. (2021), "Understanding and deciphering of social engineering attack scenarios", *Security and Privacy*, Vol. 4 No. 4, doi: 10.1002/spy2.161.

Further reading

Jones, K.S., Armstrong, M.E., Tornblad, M.K. and Siami Namin, A. (2020), "How social engineers use persuasion principles during vishing attacks", *Information and Computer Security*, Vol. 29 No. 2, pp. 314-331, doi: 10.1108/ics-07-2020-0113.

Lawson, P., Pearson, C.J., Crowson, A. and Mayhorn, C.B. (2020), "Email phishing and signal detection: how persuasion principles and personality influence response patterns and accuracy", *Applied Ergonomics*, Vol. 86, p. 103084, doi: 10.1016/j.apergo.2020.103084.

Levine, T.R. (2022), "Truth-Default theory and the psychology of lying and deception detection", *Current Opinion in Psychology*, Vol. 47, p. 101380, doi: 10.1016/j.copsyc.2022.101380.

Levine, T.R., Park, H.S. and McCornack, S.A. (1999), "Accuracy in detecting truths and lies: documenting the 'veracity effect", *Communication Monographs*, Vol. 66 No. 2, pp. 125-144, doi: 10.1080/03637759909376468.

About the authors

Keith S. Jones is a Professor of Psychological Science at Texas Tech University. He received his PhD in Experimental Psychology with an emphasis on Human Factors Psychology from the University of Cincinnati. Keith S. Jones is the corresponding author and can be contacted at: keith.s.jones@ttu.edu

McKenna K. Tornblad received her MA in Experimental Psychology with an emphasis on Human Factors Psychology from Texas Tech University. She is now a Human Factors Engineer at Pacific Science and Engineering Group.

Miriam E. Armstrong received her PhD in Experimental Psychology with an emphasis on Human Factors Psychology from Texas Tech University. She is now a staff member at the Institute for Defense Analyses.

Jinwoo Choi received his MA in Experimental Psychology with an emphasis on Human Factors Psychology from Texas Tech University. He is currently a doctoral student in Texas Tech University's Human Factors Psychology program.

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