What *does* trust *have* to do with it? Training consumers to detect digital imposter scams

Marguerite DeLiema
School of Social Work, University of Minnesota, Minneapolis, Minnesota, USA

Clifford A. Robb
Department of Consumer Science, University of Wisconsin-Madison, Madison, Wisconsin, USA, and

Stephen Wendel
Behavioral Technology, Tamarindo, Costa Rica

Abstract

Purpose – One of the insidious effects of government and business imposter scams is the potential erosion of trust among defrauded consumers. This study aims to assess the relationship between prior imposter scam victimization and present ability to discriminate between real and fake digital communications from government agencies and retail companies.

Design/methodology/approach – This paper tests whether a short, interactive training can help consumers correctly identify imposter scams without mistrusting legitimate communications. Participants were randomized into one of two control groups or to one of two training conditions: written tips on identifying digital imposter scams, or an interactive fraud detection training program. Participants were tested on their ability to correctly label emails, websites and letters as real or a scam.

Findings – This paper find that prior imposter scam victimization is not associated with greater mistrust. Compared to the control conditions, both written tips and interactive digital fraud detection training improved identification of real communications and scams; however, after a two- to three-week delay, the effect of training decreases for scam detection.

Originality/value – Results indicate that prior imposter scam victimization is not associated with mistrust, and that one-time fraud detection training improves consumers' detection of imposter scams but has limited long-term effectiveness.

Keywords Fraud, Interactive training, Inoculation theory, Randomized control trial, Phishing

Paper type Research paper

The research reported herein was pursuant to a grant from the US SSA, funded as part of the Retirement and Disability Research Consortium (RDRC). The findings and conclusions expressed are solely those of the author(s) and do not represent the views of SSA, any agency of the federal government, author affiliation(s), or the author's/authors' RDRC affiliation.

Conflicts of interest: The authors have no conflicts of interests to disclose.

Journal of Financial Crime
Vol. 32 No. 1, 2025
pp. 77-97
Emerald Publishing Limited
1359-0790
DOI 10.1108/JFC-12-2023-0314

Introduction

Technology-facilitated imposter scams are a significant form of mass marketing fraud (Federal Bureau of Investigation, 2022). In a digital imposter scam, a scammer convinces their target to transfer money by pretending to be a familiar person, a love interest, or someone affiliated with a government agency or company. Imposter scams come in many forms, but the most common are government imposter scams, such as the Social Security Administration (SSA) imposter scam, and business imposter scams, such as Amazon imposter and tech support scams. According to the Federal Trade Commission (FTC, 2023), more than 726,000 consumers filed an imposter scam complaint in 2022 and nearly one quarter reported a financial loss. In total, these scams cost consumers $2.7bn with median losses of $1,000 per consumer (FTC, 2023). These numbers underestimate the true cost of imposter scams, as only a fraction of harmed consumers report fraud to authorities (Anderson, 2021).

Although the rate of victimization following exposure to imposter scams is low (DeLiema and Witt, 2023), targeting attempts are pervasive and require constant vigilance. In a market survey of more than 36,000 American adults, 45% received imposter scam text messages, emails or calls *every day* (YouGovAmerica, 2023). During a three-month period in 2020, nearly half of US survey respondents reported that they received at least one bogus phone call from someone pretending to be an SSA official (SimplyWise, 2021).

Many consumers report significant psychological distress from incessant fraud solicitations, even in the absence of financial loss (Bailey *et al.*, 2020; Figueroa, 2019). Research indicates that fraud victimization, and even exposure to fraud, can decrease trust (Button *et al.*, 2014; Cross *et al.*, 2016; Gurun *et al.*, 2018; Harris *et al.*, 2023). Diminished trust has social and economic consequences as consumers may be more suspicious of engaging with government agencies and companies online. Therefore, it is important for organizations that are frequently impersonated by scammers, including the SSA, Internal Revenue Service (IRS), and law enforcement, to engage in efforts to restore trust among those affected by scams.

The purpose of this study is twofold. First, we assess whether victims of government and/or business imposter scams demonstrate lower levels of trust in online communications or transactions (whether that be trust in the SSA, other federal entities or businesses) compared with those who do not report prior victimization. The second aim is to assess whether an interactive online fraud training module can improve consumers' ability to accurately discriminate between real and fraudulent government and business websites and emails without eliciting undue mistrust, and whether training effectiveness persists over a moderate two- to three-week delay.

Background and literature review

Trust in e-government and e-commerce

Trust is critical for a functioning democratic society (Bjørnskov, 2012; Hardin, 2002; Zak and Knack, 2001). Unfortunately, public trust and confidence in the US government and political institutions have declined substantially since the 1960s. According to the Pew Research Center (2022), only 20% of Americans say they trust the government in Washington to do what is right "always" or "most of the time," down from 77% in 1964. This lack of trust has serious implications for public policy. Research indicates that low trust in government is associated with less support for government spending (Chanley *et al.*, 2000), reduces political participation (Bäck and Christensen, 2016; Hoogh and Marien, 2013) and undermines compliance with laws, regulations and public health policies (Hoogh and Marien, 2013; Van Dijke and Verboon, 2010).

Under the backdrop of declining social and political trust, there is a deepening reliance on online technologies to interact with government agencies (e-government) and engage in routine consumer purchasing activities (e-commerce). Today, online retail purchases account for approximately 15% of all retail sales in the USA (United States Census Bureau, 2023). According to the IRS (2023), 94% of all individual tax returns are filed electronically. Analytics.usa.gov, a website that tracks web traffic to government domains, found that there were approximately 5.2 billion visitors to all federal websites from August through October 2023 (Analytics.usa.gov).

Most e-government and e-commerce activities involve the online transfer and storage of personal and/or financial information. Prior research has shown that trust is a critical factor in determining whether consumers will share their identifying information online and engage in Web services (Kim and Peterson, 2017; McKnight *et al.*, 2002). For example, Bélanger and Carter (2008) found that trust in the internet and trust in government were positively related to the intention to use e-government. Unfortunately, the pervasiveness of government and business imposter scams, combined with high-profile data breaches and other cyber security threats, may cause mistrust and deter consumers from engaging with government and retailers online.

The impact of fraud on trust

The consequences of diminished trust have measurable impacts on financial markets and consumer confidence in the legal system. According to Goel (2021), imposter scams can undermine the government's authority to administer laws and enforce policies. Button *et al.* (2014) found that many fraud victims reported socioemotional impacts and "were more cautious about making decisions involving finance, using their credit card, and purchasing items on the internet" (p. 51). Brenner *et al.* (2020) found that fraud victims had lower confidence in financial matters compared to nonvictims. Using investment advisor and branch deposit data, Gurun *et al.* (2018) found that investors in communities more exposed to the Bernard Madoff Ponzi scheme subsequently withdrew more assets from registered investment advisers and increased deposits at banks, leading to lower returns over time. Using Gallup survey data, the authors also found that people who were more exposed to the Madoff scheme reported larger declines in confidence in the criminal justice system compared with people unaffected by the fraud. Jansen and Leukfeldt (2018) interviewed cyber fraud victims and found that for a quarter of the sample, victimization lowered trust in banks and/or online banking. Victimization also caused some participants to lose faith in their own capabilities and diminished their trust in others. Together, these findings indicate that fraud exacts a toll on trust, resulting in harmful social and economic outcomes.

Training consumers to identify digital imposter scams

One potential strategy to enhance trust in e-government and e-commerce is to teach consumers to discriminate between real and fraudulent email and website communications. Prior research focusing on fraudulent emails found evidence that such training is indeed possible and effective. Robb and Wendel (2023) demonstrated that sharing cyber security tips in a simple online tutorial may help consumers develop more confidence in their ability to flag phishing emails, thereby avoiding imposter scams.

Study hypotheses

This study sought to answer two primary research questions. First, are individuals who were previously victimized by an imposter scam less trusting than individuals who have not experienced imposter fraud? We define fraud victimization as reporting that money was

stolen in an imposter scam. We measure the effects of prior imposter fraud victimization on trust in online transactions, trust and confidence in the federal government, and trust in the SSA – the most impersonated federal agency in 2021 and 2022. Experimentally, we also assess revealed trust in government and business communications among former imposter scam victims relative to nonvictims. Our hypotheses are:

H1. Prior imposter fraud victimization is associated with lower trust in the SSA, lower trust and confidence in the US government, and lower trust in online transactions.

H2. Prior imposter fraud victimization is associated with greater mistrust of legitimate and fake communications.

The second research question is whether a short, interactive online fraud detection training program that gives users immediate feedback can help them identify digital features that signal that a website or email is a phishing attempt. Prior research indicates that fraud awareness messaging is more impactful when it is embedded within the context that scams occur (Kumaraguru *et al.*, 2007). For those previously targeted by an imposter scam, gaining knowledge on digital fraud detection may help restore trust and increase willingness to engage in online interactions with legitimate retailers and government agencies, including the SSA. We hypothesize:

H3. An interactive training will improve participants' accuracy in discerning legitimate and fake communications relative to a static training (written tips) and control conditions. This effect will be consistent for those who have and have not been victimized by an imposter scam.

H4. The interactive training does not increase mistrust of legitimate communications relative to the static training and control conditions.

H5. The interactive training does not improve fraud detection accuracy for nondigital solicitations (i.e. postal letters).

Prior research on training to improve detection of phishing attacks indicates that the effectiveness of training diminishes over time without reminders or additional training (Kumaraguru *et al.*, 2007; Zhang, 2018). Therefore, we predict that:

H6. The effect of the interactive training will diminish after a two-week delay.

Study hypotheses were preregistered and can be viewed at https://tinyurl.com/3vtefa5z

Methods

Study design

To test these hypotheses, we developed a platform in which participants first completed an online survey to assess different domains of trust and then were randomly assigned to one of four conditions (two treatment and two control). They were either trained on the characteristics of real and fraudulent digital communications (interactive training or static written tips) or were randomly assigned to read irrelevant information (control condition 1) or read general information about trust (control condition 2). Participants completed a demographic questionnaire followed by an assessment of their ability to correctly label emails, websites, and letters as real or fraudulent. Appendix 1 presents a graphical depiction of the participant experience that will be described in detail below.

Sample

Participants were recruited from Prolific, a commercial provider of online research participants who are paid to participate in surveys and experiments. Prolific provides quota-balanced samples that are nationally representative on three factors – age, sex and race/ethnicity. Research by the Prolific team (Peer *et al.*, 2022) and others (Coppock and McClellan, 2019) found that their method leads to samples that will match the target population, but care should be taken in directly generalizing to the entire US population. All participants were asked for their informed consent and the research was reviewed and approved by the Institutional Review Boards at University of Wisconsin-Madison and University of Minnesota.

A total of 5,891 people participated across the primary studies and 11 small, iterative tests of the application code, training materials and platform. Appendix 2 provides information about the prior iterations and samples. The final samples were as follows:

- *Primary sample:* 1,244 participants, quota balanced on age, gender and ethnicity to be nationally representative. They participated between July 30, 2022 and August 1, 2022. Ultimately, 1,191 people completed the study procedures and provided valid data (96% completion rate).
- *Time delay sample:* 1,251 participants, similarly quota balanced. Of those recruited, 1,213 people completed the first part of the study and provided valid data (97% completion rate) and 1,017 people returned to complete the second assessment and final questionnaire, yielding an 81% completion rate. The first round was made available to participants on August 14, 2022, and the second round was made available on September 2, 2022. The median delay period was 19 days.

Measures

Participants first completed a survey that assessed different forms of trust and frequency of online shopping. For trust in the SSA, we used a ten-item scale adapted from Carter and Bélanger (2005) and McKnight *et al.* (2002). Participants rated how much they agree with statements on a 1–7 Likert scale. Examples include, "The Social Security Administration is truthful in its dealings with me," and, "In my opinion, the Social Security Administration is trustworthy." The composite score is the mean of the responses, with higher scores indicating greater trust in the SSA.

Willingness to trust communications from government agencies may be affected by trust and confidence in the US government. We used a four-item scale from the Pew Research Center with questions such as "How much confidence do you have in the future of the USA?" The composite score is the sum of responses to the four items, in which each question is given equal weight and scaled to be from 0 (lowest confidence/trust) to 1 (highest confidence/trust).

For trust in online transactions, we used an eight-item, seven-point agreement scale also adapted from Carter and Bélanger (2005) and McKnight *et al.* (2002). Sample statements include: "Entering personal information over the internet is unsafe," and "I would hesitate to enter personal information like my name, address, and phone number on the internet." Several items were reverse scored. The composite score is the mean of the items where higher scores indicate greater trust in online transactions.

Finally, we asked participants to provide information on their frequency of online shopping using a one-item measure from Pavlou (2003). Participants complete the phrase, "I buy products online [. . .]" where "never" = 1 and "a few times a week/daily" = 6.

Randomization into experimental and control conditions

Next participants were randomly assigned to one of four conditions:

(1) Control Arm 1: Read written material about internet addiction (words = 180).

(2) Control Arm 2: Read written material on the importance of trusting the government based on a case study of the Ebola virus and disease prevention measures (words = 241).

(3) Static Training Arm 3: Read written digital fraud prevention tips on what to look for to identify a fraudulent website or phishing email (words = 435).

(4) Interactive Training Arm 4: Read written fraud prevention tips and complete an interactive training to practice labeling real and fraudulent simulated emails and websites.

Arm 4 was the primary focus of the study. After reviewing the written tips, participants in Arm 4 were directed to a custom online application where they were presented with eight mock communications from government agencies or retail businesses (four emails and four websites). The application presents a realistic interface in the form of mock websites or emails with which participants could interact. Participants could clink on links and buttons as if they were real but were informed by a pop-up message that all functions were disabled. The application logged all relevant activity to the database, including clicking on links and opening email headers. Participants in Arm 4 were asked to judge each communication as "real" or "fraudulent" and were provided with immediate feedback regarding accuracy and tips of what to look for within the context of the email or website. Figure 1 provides an example of what a user in the interactive condition would see after they selected "real" or "scam." Appendix 3 presents a table describing all mock communications used in the experiment.

Next participants completed a short demographic survey in Qualtrics. Participants in the primary sample went immediately to the assessment. Participants in the separate "time delay" sample were thanked for their time and asked to return in two weeks. The platform used to coordinate and pay participants, Prolific, experienced technical difficulties and participants were only able to return after a minimum of 16 days from the start of the study. They entered the assessment after a median of 19 days after receiving a study reminder and providing continued informed consent.

All participants were tested on their ability to correctly distinguish fraudulent appeals from real ones in the assessment phase. The experience was identical to the interactive training with three notable exceptions:

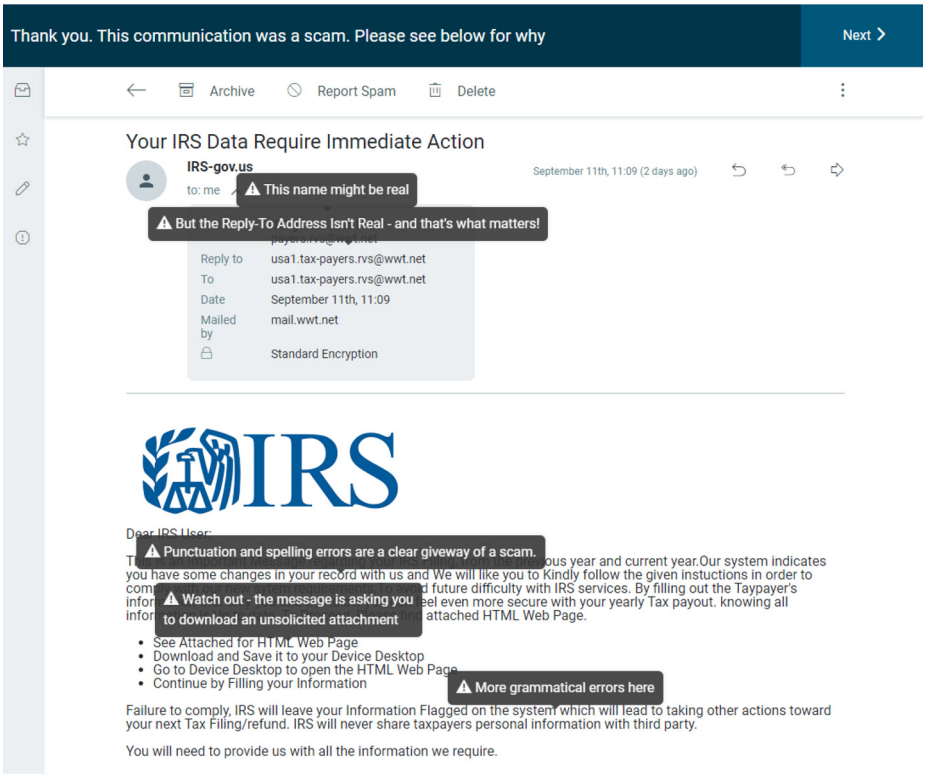
(1) Novel mock communications were presented.

(2) Participants were not informed of the correct answer after they selected "real" or "scam."

(3) The communications included two mocked-up letters – one legitimate and one a scam. Letters were included to assess the portability of the training to nondigital communication mediums.

Final survey

After completing the assessment, all participants (both immediate and time delay samples) completed a third short survey on their prior exposure and responses to various types of imposter scams and their experience with cybersecurity training. Participants answered



Notes: The top line informs the participant of the correct answer after they make a judgement on the prior screen. Tips are placed next to content that the user would use to identify this email as fraudulent
Source: Figure created by the authors

Figure 1.
An example fraudulent email used in the interactive training

whether they had previously been targeted by scammers claiming to be from the SSA, the IRS, another government agency, the company where they work, another organization or group or selling fraudulent goods or services. Participants could check any that applied. Those who indicated they had been targeted by any of these scams were asked whether they had lost money and how much (in dollars). Those who indicated "yes" were coded as victims. Following the fraud exposure and victimization questions, participants indicated whether they had previously received cybersecurity training (yes = 1).

Dependent variables in the experiment
In addition to participants' self-reported measures of trust in the SSA, trust and confidence in the US government and trust in online transactions, we calculated the following values:

- *Revealed trust.* Among the two control arms of the study where participants did not receive any training, we use the percent of real messages they correctly identified

during the assessment phase as a revealed measure of trust. Because the arms are randomly assigned, they represent a (smaller) quota-balanced sample of Americans.

- *Percent correct.* The percent of communications correctly labeled as real or a scam, with versions of this variable segmented by percent correct among business communications and government communications, as well as the percent correct among emails, websites, and letters.
- *Mistrust.* The percent of communications labeled as fraudulent, including "undue mistrust" (percent of communications that were real but labeled as fraudulent) and "rightful mistrust" (percent of communications that were fraudulent and labeled as such).

The complete surveys, training materials, communications used in the assessment and analysis code are all publicly available at https://tinyurl.com/3vtefa5z

Results

Sample characteristics

Participant characteristics are presented in Table 1. There were few statistically significant demographic differences between the quota-balanced samples [1]. In total, 49% of respondents in the primary sample were female with an average age of 41.2 years (*standard deviation* [SD] = 14.8) and 14.8 years of education (i.e., some college, SD = 2.1). In total, 48% reported a household income of less than $60,000, 26.8% reported incomes between $60,000 and $99,999, and 24.3% reported incomes of $100,000 or greater. Three-quarters of primary sample participants identified as non-Hispanic (NH) white, 11.9% as NH African American, 5.7% as NH Asian, 3.6% Hispanic, and 3.4% all other groups. Prolific balances its samples on race and does not include Hispanic or Latino/a as a separate category (considering it is an ethnicity, not a race). Because of this, the sample significantly underrepresents Hispanics.

Exposure to attempted fraud was widespread: 91.2% of participants reported being targeted by at least one type of imposter scam in the past. This is likely an underestimate given the ubiquity of attempted fraud. In the present study, 5.8% of participants (combined samples) reported victimization by an imposter scam in the past. Slightly more than one-third (35.7%) of participants reported having undergone some form of cybersecurity training.

The primary sample's mean confidence score in the US government was 0.5 (range = 0–1; SD = 0.2), trust in the SSA was 4.4 (range = 1–7; SD = 1.3), and trust in online transactions was 3.6 (range = 1–7; SD = 1.0). Revealed trust—the percent of real messages labeled as real among the untrained participants in the two control conditions in the primary sample—was 77% (SD = 19.6). Participants' overall ability to correctly label messages across all conditions was 69% (SD = 15.0), on average [2].

The impact of prior fraud victimization on self-reported trust measures

Overall, we do not find strong support for *H1* – the impact of prior imposter fraud victimization on trust in the SSA, trust and confidence in the federal government, and trust in online transactions. There were several differences in the regression results between the primary and time delay samples, so Table 2 presents findings for the primary sample only. Appendix 4 presents regression results for the time delay sample.

All else being equal, more educated adults tend to have more trust in the SSA, more trust and confidence in the US government, and more trust in online transactions. Females have

Participant characteristics	Primary sample ($N = 1,224$)		Time delay sample ($N = 1216$)		Differences between samples	
Variables	Frequency (%) or mean	SD	Frequency (%) or mean	SD	*t*-Test/chi-square	*p*-value
Years of education	14.9	2.1	14.7	2.1	*−3.07*	*0.002*
Age	41.2	14.8	37.3	12.9	*−6.5*	*<0.0001*
Sex						
Male	47.3%	–	48.7%	–		
Female	50.9%	–	48.9%	–		
Other	1.8%	–	2.4%	–	1.45	0.48
Race						
NH white	74.3%	–	72.6%	–		
Hispanic/Latino	3.6%	–	5.4%	–		
Asian	5.7%	–	5.6%	–		
Black/African American	11.9%	–	11.4%	–		
Other/mixed race	3.4%	–	2.7%	–		
Missing/prefer not to state	1.2%	–	2.2%	–	8.57	0.13
Income						
$0–$19,999	11.8%	–	12.2%	–		
$20,000–$39,999	20.1%	–	20.7%	–		
$40,000–$59,999	16.6%	–	17.9%	–		
$60,000–$79,999	14.7%	–	14.9%	–		
$80,000–$99,999	11.9%	–	10.8%	–		
$100,000–$149,999	13.9%	–	12.2%	–		
$150,000 or more	10.4%	–	10.1%	–		
Missing	0.7%	–	1.3%	–	4.27	0.75
Pretraining trust ratings and prior imposter scam experiences						
Trust and confidence in US government	0.5	0.19	0.5	0.2	1.59	0.11
Trust in SSA	4.37	1.32	4.26	1.31	*−2.05*	*0.04*
Trust in online transactions	3.61	1.01	3.52	1.03	−1.91	0.06
Exposed to imposter scam	91.2%	–	90.7%	–	0.2	0.66
Victim of imposter scam	6.1%	–	5.4%	–	0.41	0.52

Note: Significant differences between samples are *italicized*
Source: Table created by the authors

Table 1. Characteristics of the primary and time delay samples

less trust and confidence in the US government (coef = −0.05; SE = 0.01; $p < 0.001$) and less trust in online transactions (coef = −0.15; SE = 0.06; $p = 0.01$), but there is no association between sex and trust in the SSA. Nonwhite individuals report significantly higher trust in the SSA (coef = 0.23; SE = 0.09; $p = 0.01$) and higher trust and confidence in the US government (coef = 0.08; SE = 0.01; $p < 0.001$), but lower trust in online transactions (coef = −0.18; SE = 0.07; $p < 0.001$). Higher household income was associated with more trust and confidence in the US government (coef = 0.0002; SE = 0.0001; $p = 0.04$). Online shopping frequency was positively associated with all measures of trust, whereas prior cybersecurity training had no effect.

There was a nonlinear association between age and trust and confidence in the US government as well as age and trust in online transactions. We find that after roughly age 30, trust and confidence in the US government increase, all else being equal. Trust in

Table 2.
Regression results examining the correlates of trust among primary sample participants (N 1,224)

Independent variables	Trust in SSA ($n = 1{,}143$) $R^2 = 0.06$				Trust and confidence in US government ($n = 1{,}168$) $R^2 = 0.08$				Trust in online transactions ($n = 1156$) $R^2 = 0.06$			
	Coef.	SE	*t*-value	*p*-value	Coef.	SE	*t*-value	*p*-value	Coef.	SE	*t*-value	*p*-value
Intercept	3.39	0.41	8.21	<0.001	0.4	0.06	6.7	<0.001	2.14	0.31	6.79	<0.001
Prior imposter scam victim	−0.1	0.16	−0.63	0.53	−0.03	0.02	−1.09	0.27	−0.08	0.12	−0.66	0.51
Income	0.0007	0.0009	0.77	0.44	0.0002	0.0001	2.1	0.04	0.001	0.0007	1.57	0.12
Years of education	0.03	0.02	1.72	0.09	0.008	0.003	2.99	0.003	0.02	0.02	1.27	0.21
Age	−0.02	0.02	−1.43	0.15	−0.006	0.002	−2.59	0.01	0.03	0.01	2.32	0.02
Age-squared	0.0005	0.0002	2.56	0.01	0.0001	3E-05	3.1	0.002	−0.0004	0.0001	−2.42	0.02
Sex (1 = female)	−0.11	0.08	−1.4	0.16	−0.05	0.01	−4.66	<0.001	−0.15	0.06	−2.5	0.01
Nonwhite	0.23	0.09	2.56	0.01	0.08	0.01	6.47	<0.001	−0.18	0.07	−2.54	0.01
Frequency of online shopping	0.12	0.04	2.92	0.004	0.01	0.006	2.35	0.02	0.16	0.03	5.31	<0.001
Prior cyber security training	−0.10	0.08	−1.14	0.25	−0.02	0.01	−1.55	0.12	−0.10	0.06	−1.58	0.12

Source: Table created by the authors

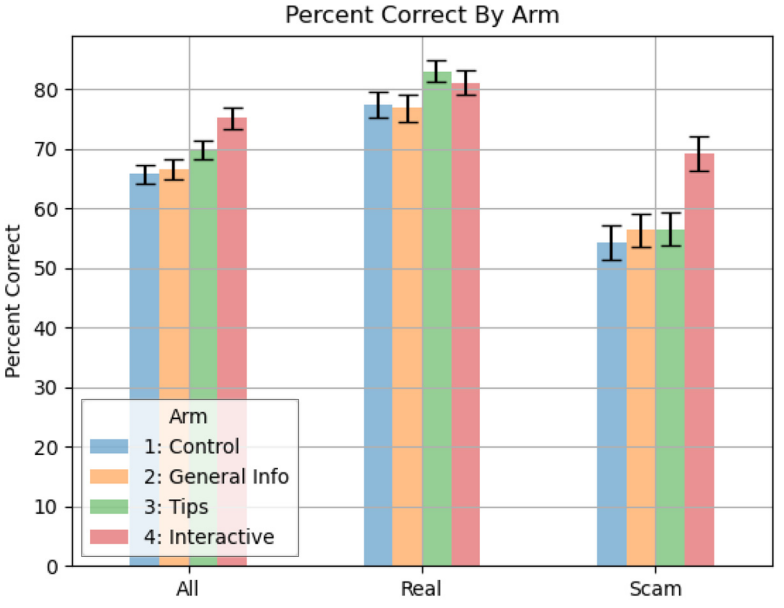
internet interactions peaks in the mid-fifties; however, the decline before and after is slight. Age was not significantly associated with trust in the SSA.

We also do not find evidence for *H2*, that prior imposter fraud victimization is related to mistrust measured using behavior-based metrics from the experiment. Prior victimization is not associated with total mistrust (total number of messages in the experiment that were flagged as fraudulent), "rightful" mistrust (total number of fraudulent messages flagged as fraudulent) and "undue" mistrust (total number of true messages flagged as fraudulent).

Effects of the interactive fraud detection training

Primary sample participants in the interactive training show a causal increase in their overall fraud-detection abilities compared with the other three conditions (Figure 2), which supports *H3*. These results are driven by an increase in correctly identifying scams. The interactive training shows a substantial 15.1 percentage point increase in fraud-detection abilities compared to Control Arm 1 (information on internet addiction), a 12.8 percentage point increase relative to Control Arm 2 (information on the importance of trusting the government) and a 12.7 percentage point increase relative to Arm 3 (noninteractive fraud detection tips). All results are statistically significant ($p < 0.001$) after Holm correction for multiple tests. The interactive training increases the ability to identify real messages by 3.8 percentage points versus Control Arm 1 and 4.3 percentage points versus Control Arm 2; it shows no statistically significant effect relative to Arm 3 and indeed shows a slightly lower, nonsignificant effect.

Importantly, we see that the experimental effect is primarily, but not exclusively, driven by teaching people to correctly label government communications as real or a scam rather than business communications (Figure 3).



Source: Figure created by the authors

Figure 2. Differences in fraud detection accuracy for real and fake communications across four conditions for the primary sample (immediate test)

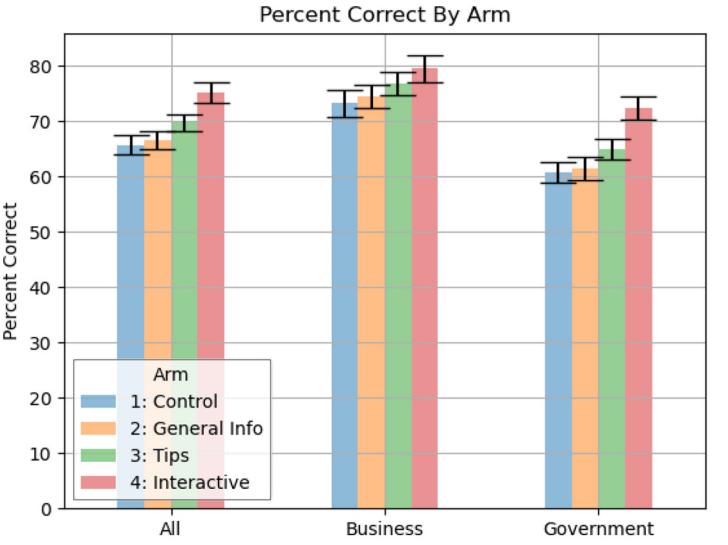


Figure 3.
Differences in fraud detection accuracy for business and government communications across four conditions for the primary sample

Source: Figure created by the authors

The interactive training also shows a causal improvement in fraud-detecting behaviors. Interactive-training participants opened 1.1 more email headers on average (from a baseline of 1.4 and out of a total maximum of four; $p < 0.001$); Arm 3 increased email open behavior by 0.53 ($p < 0.001$). The interactive training beneficially decreased the number of links clicked in the mocked-up email and website communications (coef $= -0.1$; $p = 0.01$). The noninteractive fraud detection tips had no effect on link clicks.

When we include an interaction term to test for a differential impact of the training on participants who have been victimized by an imposter scam, the baseline experimental effect remains. In other words, there is no differential impact by prior fraud victimization.

Effects of the interactive training on "undue mistrust" of legitimate communications

When designing an effective fraud awareness campaign and training consumers to detect scams, marketers and educators want to avoid increasing consumers' general levels of mistrust. In the present study, when we examine "undue mistrust" – the flagging of real messages as fraudulent – we do not see any increase in mistrust, which supports *H4*. On the contrary, both the interactive training and static fraud detection tips *decrease* "undue mistrust" (Table 3). It is notable that the static fraud detection tips showed a slightly larger

Table 3.
Effects of the training program on undue mistrust

Study condition	Coef.	SE	*t*-value	*p*-value
Intercept (Arm 1 – control information)	22.57	1.04	21.65	<0.001
Arm 2: General info on government trust	0.32	1.48	0.22	0.829
Arm 3: Static fraud detection tips	−5.69	1.48	−3.85	<0.001
Arm 4: Interactive training	−3.73	1.49	−2.51	0.012

Source: Table created by the authors

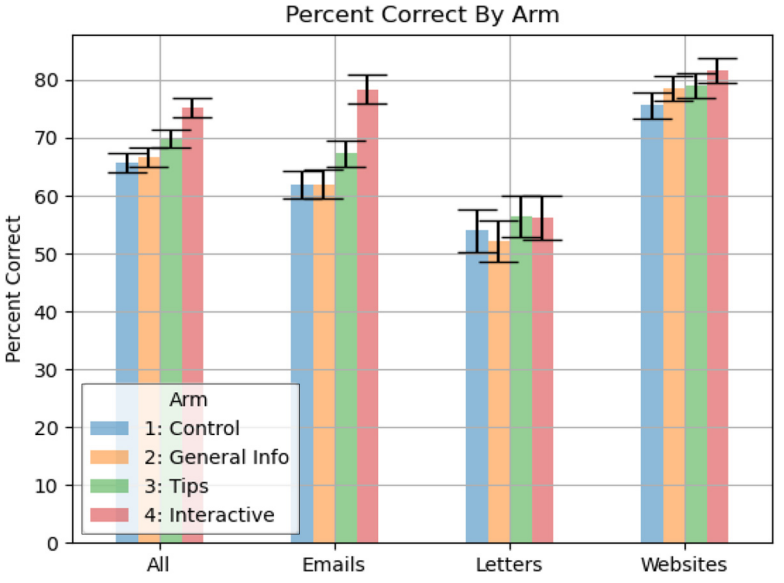
effect than the interactive training. When we examine the time-delay sample, however, only the effect of the interactive training remains after a median delay of 19 days (coef = −5.1; SE = 1.63; $p = 0.002$); the effect of the noninteractive tips fades ($p = 0.822$).

Fraud detection accuracy of different communication types

Overall, the training is more effective for emails than for websites: For emails, the interactive training resulted in an approximately 16.5 percentage point increase in accuracy over Control Arms 1 and 2 ($p < 0.001$), yet for websites, improvement in accuracy was only 6 percentage points higher than Control Arm 1 ($p < 0.01$) and 3 percentage points higher than Control Arm 2 ($p = 0.05$) (Figure 4). In the assessment, we added one real and one fraudulent letter. Arm 4 participants were no more accurate at discriminating between real and fake letters than participants in the control arms. This finding supports *H5*, which posited that digital fraud detection training does not help people identify nondigital forms of fraud. Letters also had the poorest accuracy overall (54% correct in Arm 1).

Decay in training efficacy

In support of *H6*, we find that the impact of the training on overall accuracy diminishes by half over a median delay of 19 days (Figure 5). As noted above, the interactive training boosted accuracy by 8.6 percentage points ($p < 0.001$) relative to Control Arm 2 in the immediate assessment but only increased accuracy by 4.7 percentage points ($p = 0.001$) after the delay. The decreased experimental effect is driven by a decrease in scam detection: accuracy attenuates from a 12.8 percentage point increase in the immediate assessment to a 4.9 percentage point increase ($p = 0.02$) after a delay relative to Arm 2. The effect on real-message detection is largely unchanged by time, with a boost of 4.3 percentage points in the



Source: Figure created by the authors

Figure 4. Fraud detection accuracy by type of communication across four conditions for the primary sample

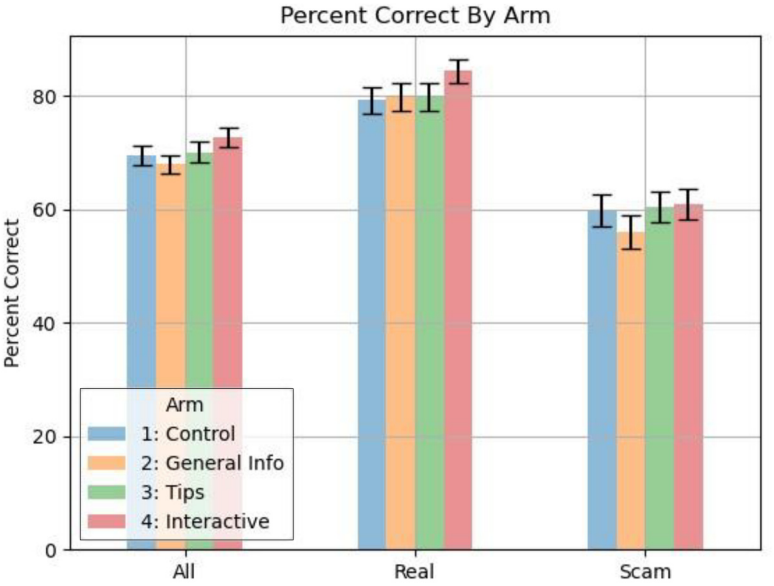


Figure 5.
Differences in fraud detection accuracy for real and scam communications across four conditions for the time delay sample

Source: Figure created by the authors

immediate assessment ($p = 0.006$) and 4.6 percentage points after the delay period ($p = 0.004$). That is indeed encouraging as the training provides an effect of boosting trust in real messages that endures over moderate time periods, even if the fraud-detection abilities fade.

After the time delay, the beneficial impact of the training is concentrated in emails and in government communications. We do not see a strong effect for websites or for business communications.

The noninteractive static training loses all its statistically significant effects in overall accuracy. Thus, although the impact of the interactive training does decay over a two- to three-week interval, the noninteractive fraud detection tips provide no enduring value in this study.

Discussion

Effects of prior imposter fraud on trust

We hypothesized that those who previously experienced an imposter scam would demonstrate lower trust (in the SSA, government in general, and online transactions). We also hypothesized that compared with nonvictims, prior fraud victims would demonstrate lower revealed trust in their judgments of the legitimacy of mock communications developed for the experiment. However, there were no significant differences between victims and nonvictims in revealed trust based on experimental findings. Overall, these results suggest that the impact of imposter scam victimization on subsequent trust may be minimal or short-lived. It is possible that the ubiquity of attempted fraud (91% of the Prolific sample) depresses trust equally across the entire population – regardless of who experienced victimization. Alternatively, imposter scam victims may have had higher baseline levels of trust than nonvictims before losing money, perhaps increasing their susceptibility. Longitudinal data are needed to determine whether there were differences in prefraud trust between victims and nonvictims.

Effects of interactive training on increasing trust in legitimate communications

Our experimental findings indicate that interactive fraud detection training can help consumers discriminate between real and fraudulent online communications and that, as hypothesized, the effect of training is no greater for those who have experienced imposter fraud than those who have not. We find that training is best at helping consumers correctly discern and label fraudulent communications as fake, rather than correctly label legitimate communications as real, although there is an effect for legitimate communications as well. This experimental finding is likely because the training focused primarily on the signs that indicate a solicitation is fake, such as alterations to a company's URL (web address) or disguising a sender's email address. There were relatively few tips on the indicators that a message is legitimate. Future training programs may seek to incorporate additional tips on the indicators of authenticity.

We find that the interactive training is more effective for email fraud detection compared with websites. The interactive training incorporated images to demonstrate actions like expanding email headers to examine the "reply to" email address. To detect a fraudulent website, there are no expandable headers for users to investigate, and it is possible that the written tips on inspecting the URL are sufficient. Moreover, the link "hover" function is not available on most touchscreen devices, so phone and tablet users cannot preview where links direct Web traffic.

It is noteworthy that the noninteractive tips provide a lesser but statistically significant benefit and require no special platform for training. Unfortunately, the effects of the written tips completely dissipated after the delay period.

Results show a stronger training effect for government communications than for business communications – overall accuracy was greater. One explanation is that consumers may have more existing expertise in assessing the legitimacy of business communications than they do for communications from government agencies, given that government communications are typically less frequent. Also, when inspecting business emails and websites, consumers may pay attention to other salient aspects that were not highlighted in the training, such as consumer reviews and whether they have an existing relationship with the company. This effect may also be caused by the quality of the training mock-ups – "government" training materials could have been systematically more effective than the "business" training materials.

Participants assigned to the interactive and written training conditions were no more accurate than those assigned to the control conditions at labeling the real and fake letters. In other words, the training did not port to novel, nondigital forms of communication. Participants demonstrated relatively poorer fraud detection accuracy for letters, suggesting that the SSA and other organizations should carefully consider the use of the mail when corresponding with consumers.

Importantly, we found that both interactive and static training does not cause an increase in undue mistrust, whereby participants incorrectly label real messages as scams after being forewarned. In other cybersecurity training research, Kumaraguru *et al.* (2009) similarly found that training does not increase the likelihood that participants falsely flag legitimate emails as phishing attacks. Similarly, Burke *et al.* (2022) found only slight reductions in willingness to invest in legitimate investment opportunities among participants trained on investment scam tactics, suggesting that forewarnings can improve scam detection without substantially decreasing trust.

Like prior studies, we find that the fraud-detection effect diminishes with time. Zhang (2018) found that technical knowledge of phishing attacks wore off within 15 days following the training program, which is a similar period to what we observed. Burke and colleagues

(2022) saw the effect of video and text-based investment scam trainings disappear at six months; however, the effect of the training persisted for participants who received a reminder after three months. In the present study, although training effects decay for scam detection, the accurate detection of legitimate messages did not substantially decrease. Future research should test whether delivering a "booster" training or other reminder message restores the effect.

Limitations

This study could include only 12 mock communications based on real and fake communications. These are only one small part of the evolving body of techniques that scammers use to deceive consumers. Effects may differ if the training was delivered outside of a research context, used mock emails or websites from different government agencies or more obscure companies, or if it was provided immediately after a participant experienced attempted fraud. Future studies should assess trust immediately following victimization.

Another limitation is that fraud exposure is so widespread (91%) that we cannot effectively assess the impact of mere exposure on self-reported or revealed trust. Among the nearly 6% of participants who indicated victimization, we did not ask follow-up questions asking them to rate the psychological and financial impact. It is possible that trust levels are only affected for those with severe imposter scam experiences.

Conclusion

This research began with the concern that widespread business and government imposter scams would undermine the trust of citizens in their government, and especially in frequently impersonated agencies like the SSA. Contrary to our hypotheses, self-reported trust and behavior-based trust measures were not lower among those who lost money in an imposter scam. This is good news for public trust and engagement in e-government; however, even if imposter scams do not appear to be undermining public trust, they consume law enforcement resources and redirect agency efforts toward educating the public about fraud rather than administering vital programs.

Experimental research into fraud susceptibility and techniques to shore up consumer defenses are still quite nascent; however, a growing body of research is pointing to the effectiveness of fraud education as a protective mechanism. Our findings indicate that government agencies and online retailers can help by incorporating fraud awareness tips in their digital communications. Future research is needed to develop strategies to improve knowledge retention, as effectiveness of fraud detection training decays over time.

Notes

1. Participants in the time delay sample were slightly younger than primary sample participants (37.3 years) and had fewer years of education (14.7 years). Time delay participants also reported slightly lower trust in the SSA. No other characteristics were statistically different between samples.
2. Baseline trust metrics are nearly identical for the primary and time delay samples. There were differences in the ability to correctly label messages, which is also to be expected because the time between training and assessment was longer (mean = 19 days) for the time delay sample.

References

Anderson, K.B. (2021), "To whom do victims of mass-market consumer fraud complain?", available at SSRN 3852323.

Bäck, M. and Christensen, H.S. (2016), "When trust matters—a multilevel analysis of the effect of generalized trust on political participation in 25 European democracies", *Journal of Civil Society*, Vol. 12 No. 2, pp. 178-197.

Bailey, J., Taylor, L., Kingston, P. and Watts, G. (2020), "Older adults and 'scams': evidence from the mass observation archive", *Journal of Adult Protection*, Vol. 23 No. 1, pp. 57-69.

BéLanger, F. and Carter, L. (2008), "Trust and risk in e-government adoption", *The Journal of Strategic Information Systems*, Vol. 17 No. 2, pp. 165-176.

Bjørnskov, C. (2012), "How does social trust affect economic growth?", *Southern Economic Journal*, Vol. 78 No. 4, pp. 1346-1368.

Brenner, L., Meyll, T., Stolper, O. and Walter, A. (2020), "Consumer fraud victimization and financial well-being", *Journal of Economic Psychology*, Vol. 76, p. 102243.

Burke, J., Kieffer, C., Mottola, G. and Perez-Arce, F. (2022), "Can educational interventions reduce susceptibility to financial fraud?", *Journal of Economic Behavior and Organization*, Vol. 198, pp. 250-266.

Button, M., Lewis, C. and Tapley, J. (2014), "Not a victimless crime: the impact of fraud on individual victims and their families", *Security Journal*, Vol. 27 No. 1, pp. 36-54.

Carter, L. and Bélanger, F. (2005), "The utilization of e-government services: citizen trust, innovation and acceptance factors", *Information Systems Journal*, Vol. 15 No. 1, pp. 5-25.

Chanley, V.A., Rudolph, T.J. and Rahn, W.M. (2000), "The origins and consequences of public trust in government: a time series analysis", *Public Opinion Quarterly*, Vol. 64 No. 3, pp. 239-256.

Coppock, A. and McClellan, O. (2019), "Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents", *Research and Politics*, Vol. 6 No. 1, pp. 1-14.

Cross, C., Richards, K. and Smith, R.G. (2016), "The reporting experiences and support needs of victims of online fraud", *Trends and Issues in Crime and Criminal Justice*, Vol. 518, pp. 1-14.

DeLiema, M. and Witt, P. (2023), "Profiling consumers who reported mass marketing scams: demographic characteristics and emotional sentiments associated with victimization", *Security Journal*, pp. 1-44.

Federal Bureau of Investigation (2022), "Federal Bureau of Investigation Internet Crime Report 2022", available at: www.ic3.gov/Media/PDF/AnnualReport/2022_IC3Report.pdf

Federal Trade Commission (2023), "Consumer Sentinel Network Data Book 2022", available at: www.ftc.gov/reports/consumer-sentinel-network-data-book-2022

Figueroa, C. (2019), "The pallone-thune 'TRACED act': expanding consumer protection in the fight against robocalls", *Loyola Consumer Law Review*, Vol. 32 No. 2, pp. 318-331.

Goel, R.K. (2021), "Masquerading the government: drivers of government impersonation fraud", *Public Finance Review*, Vol. 49 No. 4, pp. 548-572.

Gurun, U.G., Stoffman, N. and Yonker, S.E. (2018), "Trust busting: the effect of fraud on investor behavior", *The Review of Financial Studies*, Vol. 31 No. 4, pp. 1341-1376.

Hardin, R. (2002), *Trust and Trustworthiness*, Russell Sage Foundation, New York, NY.

Harris, E., Petrovits, C. and Yetman, M. (2023), "Spreading the news: donor response to disclosures about nonprofit fraud and efforts to rebuild trust", *Nonprofit and Voluntary Sector Quarterly*.

Hoogh, M. and Marien, S. (2013), "A comparative analysis of the relation between political trust and forms of political participation in Europe", *European Societies*, Vol. 15 No. 1, pp. 131-152.

Internal Revenue Service (2023), "Returns filed, taxes collected and refunds issued", available at: www.irs.gov/statistics/returns-filed-taxes-collected-and-refunds-issued

Jansen, J. and Leukfeldt, R. (2018), "Coping with cybercrime victimization: an exploratory study into impact and change", *Journal of Qualitative Criminal Justice and Criminology*, Vol. 6 No. 2, pp. 205-228.

Kim, Y. and Peterson, R.A. (2017), "A meta-analysis of online trust relationships in e-commerce", *Journal of Interactive Marketing*, Vol. 38 No. 1, pp. 44-54.

Kumaraguru, P., Rhee, Y., Sheng, S., Hasan, S., Acquisti, A., Cranor, L.F. and Hong, J. (2007), "Getting users to pay attention to anti-phishing education: evaluation of retention and transfer", Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit, pp. 70-81.

Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M.A. and Pham, T. (2009), "School of phish: a real-world evaluation of anti-phishing training", Proceedings of the 5th Symposium on Usable Privacy and Security, pp. 1-12.

McKnight, D.H., Choudhury, V. and Kacmar, C. (2002), "Developing and validating trust measures for e-commerce: an integrative typology", *Information Systems Research*, Vol. 13 No. 3, pp. 334-359.

Pavlou, P.A. (2003), "Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model", *International Journal of Electronic Commerce*, Vol. 7 No. 3, pp. 101-134.

Peer, E., Rothschild, D., Gorden, A., Evernden, Z. and Damer, E. (2022), "Data quality of platforms and panels for online behavioral research", Behavior Research Methods, Vol. 54 No. 4, pp. 1643-1662, doi: 10.3758/s13428-021-01694-3. 34590289.

Pew Research Center (2022), "Public trust in government: 1958-2022", available at: www.pewresearch.org/politics/2022/06/06/public-trust-in-government-1958-2022/

Robb, C.A. and Wendel, S. (2023), "Who can you trust? Assessing vulnerability to digital imposter scams", *Journal of Consumer Policy*, Vol. 46 No. 1, pp. 27-51.

SimplyWise (2021), "SimplyWise retirement confidence index January 2021", available at: www.simplywise.com/blog/retirement-confidence-index/

United States Census Bureau (2023), "Quarterly retail e-commerce sales 2nd quarter 2023", available at: www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf

Van Dijke, M. and Verboon, P. (2010), "Trust in authorities as a boundary condition to procedural fairness effects on tax compliance", *Journal of Economic Psychology*, Vol. 31 No. 1, pp. 80-91.

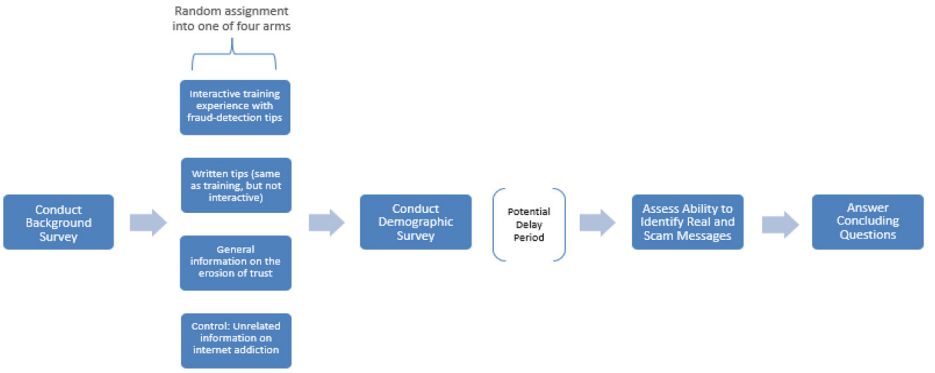
YouGovAmerica (2023), "How often would you say that you receive scam texts, emails, or calls from scammers purporting to be someone they are not?", available at: https://today.yougov.com/topics/technology/survey-results/daily/2023/03/06/8073a/1

Zak, P.J. and Knack, S. (2001), "Trust and growth", *The Economic Journal*, Vol. 111 No. 470, pp. 295-321.

Zhang, T. (2018), "Knowledge expiration in security awareness training", Annual ADFSL Conference on Digital Forensics, Security and Law, Vol. 2, pp. 197-212.

Corresponding author
Marguerite DeLiema can be contacted at: mdeliema@umn.edu

Appendix 1



Source: Figure created by the authors

Figure A1. Visual depiction of the study procedures

Appendix 2. Description of prior samples and iterative tests of the materials and study design

Before the main study, 1,128 participants were invited to participate to test the application code. These participants came from Prolific's convenience participant pool and were not recruited to be nationally representative. Next, a quota-based test sample of 1,226 participants from Prolific were recruited to assess the full study flow from July 9, 2022 to July 13, 2022. Of these participants, 1,192 people completed the survey and provided valid data (97% completion rate). Each iteration excluded participants in its sample from prior studies. In our initial analysis of the results from this sample, we determined that several of the test communications had misleading elements that participants were using to judge legitimacy, rather than the tips they were trained on in the experimental conditions. We corrected these elements before recruiting the two final samples described in the main body of the article.

Appendix 3

Table A1. Mock digital communications used in the interactive training condition (Arm 4) and final knowledge test (all participants)

Type	Training or test	Legitimacy	Imposter type	Short name	Description
Email	Training	Real	Government	SSA scam warning	An email providing information on the warning signs of SSA imposter scams
Web	Training	Real	Business	Amazon homepage	The real Amazon.com homepage
Email	Training	Scam	Government	IRS email	Fake email from the IRS requesting that the taxpayer download an attachment
Web	Training	Real	Government	IRS online account	The sign-in page to access online IRS accounts
Web	Training	Real	Business	Apple ID sign-in	Apple sign-in page for Apple account management
Web	Training	Scam	Business	Amazon product	A fake product page selling home security cameras
Email	Training	Real	Business	McAfee email	A monthly security activity report from McAfee ID Theft Protection
Email	Training	Scam	Government	SSA security email	An email announcing a new security alert feature from SSA that requires sign-up
Website	Test	Real	Business	Walmart product	A real product page from Walmart.com selling security cameras
Email	Test	Scam	Business	Red Cross	An email asking for donations to a fraudulent Red Cross site
Website	Test	Real	Government	SSA HP	The SSA's actual homepage
Website	Test	Scam	Government	mySSA	A modified version of the mySSA login page
Email	Test	Real	Business	Amazon mark delivery	An email informing the recipient of a delay in the delivery of an Amazon product
Website	Test	Scam	Business/ government	SSA Facebook	A Facebook page for a fake Social Security Administration account
Email	Test	Scam	Government	SSA opt out	A modified version of an SSA email allowing people to opt out of future SSA emails
Email	Test	Real	Government	Replacement card	An email from the SSA instructing recipients on how to get a new SSA card
Letter	Test	Real	Government	Medicare review	A notice from the SSA about an upcoming call from the Medicare office
Letter	Test	Scam	Government	Benefits suspension	A letter threatening the suspension of SSA benefits

Source: Table created by the authors

Appendix 4

Table A2. Regression results examining the correlates of pretraining measures of trust among time delay sample participants (N 1,216)

	Trust in SSA ($n = 1{,}138$) $R^2 = 0.02$				Trust and confidence in US government ($n = 1{,}158$) $R^2 = 0.06$				Trust in online transactions ($n = 1{,}138$) $R^2 = 0.06$			
Independent variables	Coef.	SE	t-value	p-value	Coef.	SE	t-value	p-value	Coef.	SE	t-Value	p-value
Intercept	3.53	0.41	8.62	<0.001	0.23	0.06	3.74	<0.001	2.33	0.32	7.39	<0.001
Prior imposter scam victim	0.13	0.16	0.83	0.410	−0.01	0.02	−0.37	0.710	−0.15	0.13	−1.2	0.230
Income	0.002	0.001	1.9	0.060	0.0003	0.0001	2.26	0.020	9E-04	0.0007	1.41	0.160
Years of education	0.03	0.02	1.53	0.130	0.02	0.003	5.81	<0.001	0.02	0.02	1.53	0.130
Age	−0.008	0.02	−0.46	0.650	*0.0004*	*0.003*	*0.17*	*0.870*	0.03	0.01	2.18	0.030
Age-squared	*0.0002*	*0.0002*	*1.1*	*0.270*	*−1E-05*	*3E-05*	*−0.27*	*0.790*	−0	0.0002	−2.34	0.020
Sex (1 = female)	−0.05	0.08	−0.69	0.490	−0.02	0.01	−2.13	0.030	−0.23	0.06	−3.95	<0.001
Nonwhite	*0.05*	*0.09*	*0.54*	*0.590*	0.04	0.01	3.21	0.001	−0.26	0.07	−3.79	<0.001
Frequency of online shopping	*0.04*	*0.04*	*0.88*	*0.380*	*0.005*	*0.006*	*0.88*	*0.380*	0.11	0.03	3.49	<0.001
Prior cyber security training	−0.11	0.08	−1.37	0.170	−0.01	0.01	−0.88	0.380	−0.01	0.06	−0.2	0.840

Note: Differences in effect sizes and statistical significance from the primary sample are given in italic
Source: Table created by the authors