Center for Financial Security
UNIVERSITY OF WISCONSIN-MADISON
Retirement & Disability Research Center

Clifford Robb
University of Wisconsin-Madison
carobb@wisc.edu

Stephen Wendel
Behavioral Technology
steve@behavioraltechnology.co

Assessing Vulnerability to Social Security Scams

Center for Financial Security

University of
Wisconsin-Madison

1300 Linden Drive
Madison, WI 53706

608-890-0229
cfs@mailplus.wisc.edu
cfs.wisc.edu

The research reported herein was performed pursuant to a grant from the U.S. Social Security Administration (SSA) funded as part of the Retirement and Disability Consortium. The opinions and conclusions expressed are solely those of the author(s) and do not represent the opinions or policy of SSA or any agency of the Federal Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation or favoring by the United States Government or any agency thereof.

1. Abstract

Over the last few years, Social Security scams have become one of the most common forms of government imposter fraud. These scams cost innocent people in the United States millions of dollars each year and undercut the ability of the Social Security Administration to contact and interact with citizens about their benefits. This report presents research into how to help individuals discriminate between scams and real appeals from the Social Security Administration. On a nationally representative sample of United States residents, the authors randomly assign participants to one of four training programs: from general tips about scams to a targeted experiential learning program inspired by *inoculation theory*. There is strong evidence that the inoculation process successfully and significantly increases fraud detection without decreasing trust in real communications. It provides protection against both SSA and non-SSA scams, such as Amazon imposter scams. The impact, however, is specific to the mode of communication (email versus letter or SMS) and decays over time; training programs should be targeted accordingly. This study suggests that a low cost, four and a half minute training can help individuals fight fraud, and such training should be examined for further refinement and potentially for broad deployment.

Keywords: Scam Identification; Behavioral Science; Digital Fraud; Inoculation Theory; Randomized Control Trial

JEL Classification: D91; P46

2. Background

In 2020, the Social Security Administration (SSA) received complaints from more than 700,000 people reporting that they were targeted by an SSA imposter scam (Skiba 2021). Such imposter scams are known to have cost innocent people in the United States millions of dollars each year (Fletcher 2019). This fraud occurs despite widespread information: warnings are frequently provided on local and national news (e.g., KLEW 2020) and by organizations from the American Association of Retired Persons to the Consumer Protection Bureau (AARP 2019; Scheithe 2020). Guidance on recognizing and countering these scams is also available for individuals through the Social Security Information website and elsewhere[1]. Yet, the scams, and the losses, continue.

Research into scams falls into three major categories: research documenting its prevalence and permutations, research on who is susceptible to fraud, and research on how to counter it. The following sections address each in turn.

2.1 The Anatomy of an Imposter Scam

In an imposter scam, the scammer impersonates another person or organization to gain trust, with the ultimate goal of extracting personal information or money. These imposter scams can occur in person, over the phone, or via digital means such as email or text. Digital imposter scams often start with phishing attacks: in which a technical trick is used to aid the deception – such as a spoofed email address, image, or entire website – "luring" people into trusting the scammer. Conceptually, one can think of "imposter scam" as referring to the psychological means of gaining the victim's trust (posing as a trusted individual or company), and "phishing" as referring to the technical means of accomplishing it (employing a fake email address that looks like it came from the trusted company).[2]

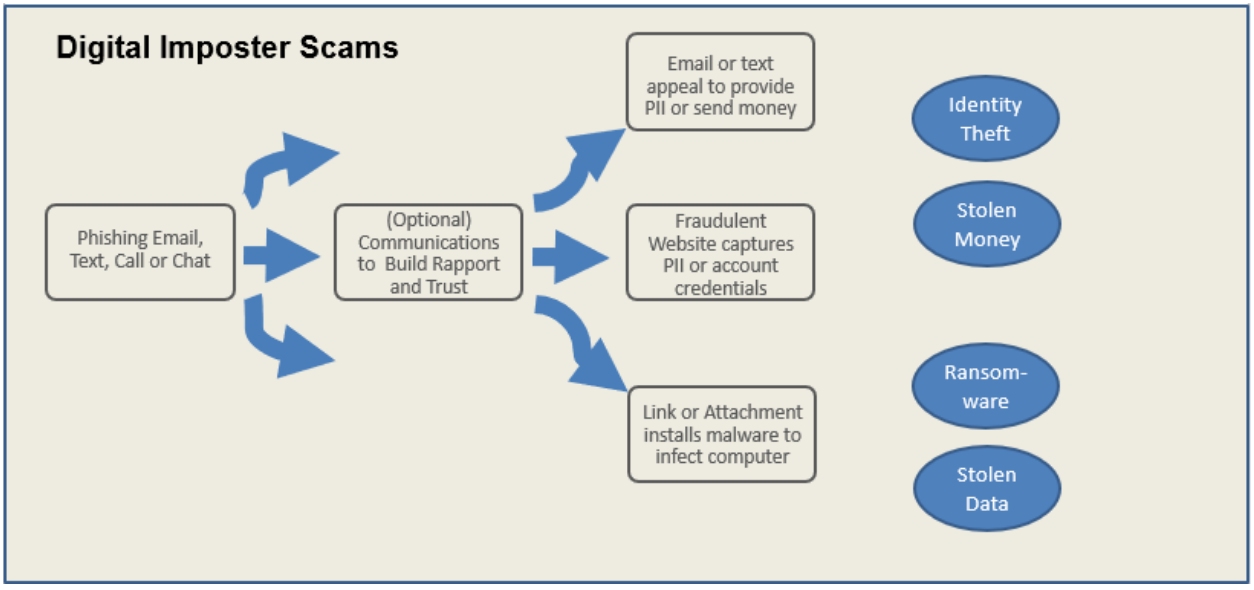
---

[1] See https://www.ssa.gov/ and https://www.ssa.gov/fraud/ for more information on combating fraud as well as the benefits of fraud reporting

[2] These are our definitions based on common usage in the field. There are no simple consensus definitions, in part because the terms are prevalent in different communities: phishing is used in cyber security, imposter scam is used in consumer protection and anti-fraud law enforcement. There is significant overlap between the concepts, but not complete: there are in-person imposter scams in which no digital deception (phishing) is used and theoretically a phishing attack does not have to impersonate a trusted party, though they usually do.

This research focuses on what the authors define as *digital imposter scam*s, in which the scammer uses digital means to impersonate a trusted organization or person. This process starts with a fraudulent email (*phishing)* or SMS (sometimes also referred to as *smishing).* The victim then either directly provides targeted information such as a social security number or interacts further with the scammers via email, web, or phone before being asked to provide personal information or money. A phishing attack in which the scammer impersonates a trusted third party could also be used to install malware on the victim's computer such as ransomware, which locks the person's computer until a ransom is paid. Figure 1 shows how the process works.

Figure 1: The Anatomy of a Digital Imposter Scam, showing how phishing/smishing/etc. is used to gain trust and extract money or information from targets.



In practice, Social Security Administration imposter scams appear to focus on the direct extraction of information or money, rather than on installing ransomware (Waggoner 2020). Figures 2 and 3 provide two examples of actual messages scammers have used to initiate the process.

Figure 3: Example of a Fraudulent SMS

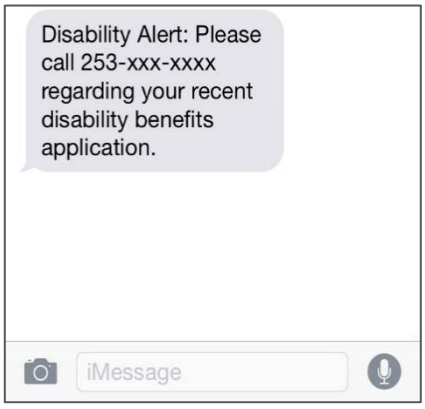


Figure 2: Example of an Email Used by Social Security Imposter Scammers

2.2 The Prevalence of Imposter Scams, and Digital Imposter Scams in Particular

The most detailed and extensive dataset on fraud in the US comes from the Federal Trade Commission's Sentinel Network. The Sentinel Network gathers reports of fraud data directly from the public and from third party sources such as the Better Business Bureau and state law enforcement agencies (FTC 2021). They provide both public access to summary statistics and private access to selected researchers.

In their 2020 data, identity theft is the most common type of report. Imposter scams (digital or otherwise), however, are the most common form of fraud and second only to identity theft in total reports. Imposter scams account for nearly 36 percent of all dollars reported lost to fraud in 2020 (FTC 2021).

In terms of the means of contact, phone calls are the most common way that fraud is initiated. However, *digital* contact by scammers (whether for an impostor scam or otherwise) is actually more common than phone calls and result in greater aggregate losses: from fraudulent websites, to emails, to social media and text messages (FTC 2021). Similarly, data from the FBI shows that phishing attacks (often overlapping with imposter scams, as described above) were the most common type of cybercrime in 2020 (FBI 2021).

When one examines Social Security scams in particular, robo-calls impersonating the SSA have been prevalent since 2018. In these calls, a recorded voice claims to be from the SSA and

threatens the callee with a range of penalties if they do not respond (Leach 2018).[3] These calls do not appear to be very effective though, as only a tiny fraction of people who report them actually fall for them (Better Business Bureau, 2021), and scammers have branched out into email as well (Waggoner 2002). These calls and emails are only the latest form of a much older phenomenon. Impostor fraud comes in waves – the 'popularity' of different types of impostors change over time; currently, SSA and business impostor scams are common; previously, it was IRS scams (Fletcher 2019).

The FTC's data on the prevalence of these scams suffers from two primary limitations, however. The first is that the ability to gather data from other sources is incomplete. The AARP, for example, using data from the Social Security Administration's Office of the Inspector General, states that there were 718,342 reports of Social Security imposter scams in 2020 (Skiba 2021). For the same period, the FTC shows only 498,278 reports across all types of imposter scams (FTC 2021).

The second limitation is more fundamental and affects both the FTC and all other fraud-report collectors: underreporting. These numbers are likely to be a vast understatement (FINRA 2013); prior research on online fraud has shown that only a fraction of targets and victims report it to authorities, and one can reasonably assume the same for SSA scams. Surveying individuals provides a more direct means of estimating these attempts than relying on self-reports. In a survey of adults in the United States, for example, the personal finance app and website SimpleWise estimated that *46 percent of all US adults* had experienced a Social Security scam attempt just in the final three months of 2020 (SimpleWise 2021).

2.3 Susceptibility to imposter scams

Research specifically on the psychology and dynamics of Social Security scams is limited, but there is a larger literature on online fraud (e.g., Chen, Beaudoin, and Hong 2017) and general financial fraud (e.g., Burnes et al. 2017) that one can draw from. This broader fraud literature can help us understand fraud and who is susceptible to it.

[3] Links to a sample recording of one of these calls can be found on at https://www.consumer.ftc.gov/blog/2018/12/what-social-security-scam-sounds

Old age is commonly believed to be a factor (e.g., Whitty 2019), but significant contrary evidence exists – showing that younger people are more likely victims than the elderly (e.g., Titus et al. 1995; Muscat et al. 2002). There may be personality characteristics that predispose an individual to falling prey to fraud – such as comfort with financial risk (Van Wyk and Benson 1997) and high impulsivity or temporal myopia (Holtfreter et al. 2008; Whitty 2019), but the area is under-researched, especially in the intersection of personality and different types of fraud. Commercial "phishing IQ tests" do exist, but they don't appear to meaningfully identify susceptibility (Anandpara et al. 2007).

As with prevalence research, there is a major challenge with work on susceptibility: there is significant and unknown information missing in the data. Many individuals simply do not know they were defrauded, and others know they were, but do not report it. Victim-blaming makes it even more difficult to accurately assess the problem. To accurately assess susceptibility, one should effectively divide the number of "successful" frauds perpetrated on a particular group by the number of attempts at fraud against them. In the existing data, however, three factors comingle:

1. Frequency by which groups are *targeted,*
2. Responsiveness of each group, and
3. Reporting rate of each group.

For example, older people in the United States are more likely to report *attempted* fraud (which may or may not mean they are more often targeted) and are less likely to report *losing money* than younger people (FTC 2021) – which may lead to some of the contradictory results researchers have found on the role of age and fraud susceptibility.

The ideal approach is to study susceptibility is not *observational* (using existing reports), but *experimental.* For example, to study fraud susceptibility, ideally one would mimic it and study who responds. The approach taken in this research measures who is susceptible within a national sample of United States residents. The measurement is based on actual behavior in a fraud experiment instead of relying on self-reports.

2.4 An Inoculation Approach

This study applies an old theory to a new problem: inoculation theory (McGuire 1961, 1964; for more recent summaries, see Compton 2013 and Banas and Rains 2010). Inoculation theory posits that exposing people to a weakened version of a dangerous appeal can help them learn to identify and resist such appeals over time. The concept has a long history of application to types of persuasion and, more generally, to threats in the future. As explained by Compton et al. in their 2016 review:

> "By exposing individuals to a persuasive message that contains weakened arguments... individuals would develop resistance against stronger, future persuasive attacks."
>
> - Compton et al. (2016)

The approach, also known as *pre-bunking*, has successfully been used to combat fake news (Roozenbeek and van der Linden 2019) and against recruitment by extremists (Saleh et al. 2021; Wendel 2020). These two specific papers are the inspiration for this work: by giving people practical experience in resisting a weakened version of impostor scams, can one help prepare them to resist the full version later?

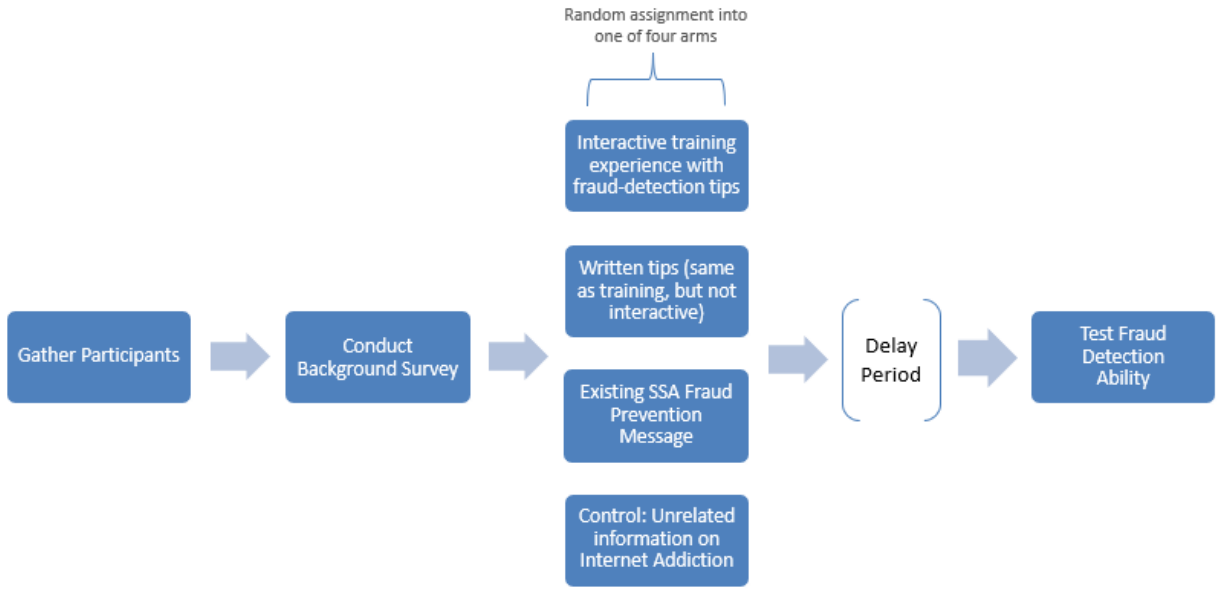
3. Research Methodology

This study tests the effectiveness of inoculating people against social security scams. It is structured as a series of randomized control trials in which participants are trained and subsequently tested for their ability to correctly identify real and fraudulent communications. The core research design is as follows:

1. Participants are directed to a website in which they complete a series of questions that cover demographic information, a measure of generalized trust, and prior experience with scams.
2. Each participant is randomly assigned to either:
    a. Interact with potential scammers via simulated email exchanges and learn how they operate (i.e., "pre-bunking");
    b. Receive written instructions on what to look for, covering the same topics and techniques as the previous arm without the interactive experience;
    c. Receive existing materials used by the Social Security Administration to inform and warn people against scams; or
    d. Receive an innocuous control condition (reading material about Internet addiction).
3. After a delay period, participants were tested on their ability to correctly distinguish fake appeals from real communications from the Social Security Administration and third parties such as the Red Cross and Amazon.com. Twelve communications in total were tested: eight emails, two letters, and two text messages.

Figure 4 illustrates the structure of the experiment.

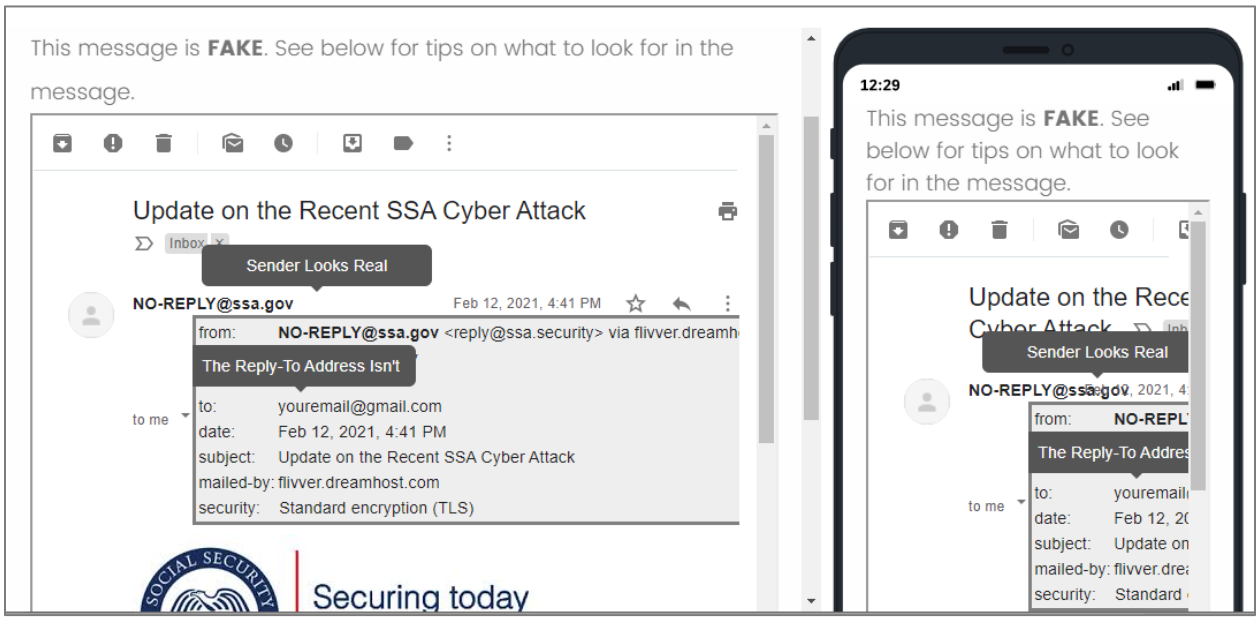
Figure 4: Experimental Design



3.1 Participant Experience: Training

The interactive training or *inoculation* arm consisted of six mocked up communications. In each case, participants were shown the communication first and asked to judge whether it was real or fake. Then, they were told the true status of the communication and provided with a set of tips overlayed and contextualized within the communication, showing them what to look for. Figure 5 shows that experience with the tips they would see after they have made their selection.

Figure 5: An Example of the Interactive Training, Showing Users What to Look for in a Communication, After They Have Guessed Whether It is Real or Fake.



Three of the messages are fake: designed to deceive the individual with spoofed email addresses and URLs, appeals for money, and so forth. Three of the messages are real: drawn directly from the Social Security Administration's actual communications. The first five communications are emails and the last one is a (real) letter from the Social Security Administration.

Individuals can interact with the emails as they would in a real environment. They can hover over links and images to see where they lead. They can open or close the email headers to learn more about the routing of the message. They can also click on links in the email, however, the traffic is intercepted, and a notice is provided informing them of the same. The details of each communication can be found in Appendix A.

The non-interactive arms of the study have a much simpler design. In each those arms, participants are presented with static information. The control arm presents reading material of similar length about Internet Addiction. The next arm shows the written tips provided at the start of the interactive experience, without any interaction. The third arm replicates existing materials that the Social Security Administration emails to individuals to warn them about scams.

The initial demographic survey and training materials were programmed in Qualtrics, augmented with Javascript, HTML, and CSS formatting to better resemble a real, interactive email environment.

3.2 Participant Experience: Test

The fraud detection test was implemented in two ways:

1. A Qualtrics platform, which is visually identical to the interactive training module, but does not provide users with the 'correct answer' to each message.
2. A real email platform (Rainloop) and a phishing-security tool (GoPhish), which looks visually dissimilar to the training module, but has the same functionality.

The two platforms serve an important research purpose: they help us understand whether the visual design of the training matters and whether the lessons generalize to other mail applications.[4] Since Rainloop is likely unfamiliar to many researchers, it warrants a brief exploration into how it works and the participant experience. Rainloop is an open-source email client and a competitor to Gmail's web interface. It looks and behaves much like Gmail or any other modern web-based email program: with a pane showing the list of available messages, and a pane in which the user can read and interact with the messages themselves. In our study, participants using Rainloop were first provided with instructions on how to operate it and asked to flag any message they believed were fraudulent as "spam". Real messages were to be left as is. Most importantly, all actions were anonymously logged to a database for analysis by researchers.

Behind the scenes, Rainloop was powered by a fully functional phishing-security environment called GoPhish. Specifically, the researchers established an email server (in Amazon's cloud) and automatically generated an email account for each person participating in the study. While the participant was completing the study's training module, Qualtrics sent an

[4] They also offer options for researchers who may want to use the freely provided code from the project's GitHub site: https://github.com/sawendel/ssascams. The Qualtrics platform is a lightweight and familiar environment to conduct additional research on these topics. The Rainloop environment requires more setup and technical sophistication, but provides a more customizable, real email experience.

API call to our software, triggering GoPhish to automatically generate the eight emails for the participant and send them to the email account created for them. More information on the technical architecture can be found in Appendix B.

The content of the test was the same for both the Rainloop-based and Qualtrics-based testing platforms. In each case, participants first interacted with eight email messages: three of them real, and five fake. Four of the messages were purportedly from the Social Security Administration, one from a disability benefits attorney, two from Amazon, and one from the Red Cross. The non-Social Security messages were included as an additional test of the generalizability of the training. A key question in the field is whether fraud detection training provides specific assistance against that fraud campaign or can benefit individuals more broadly against other types of fraud.

After the email interactions, for both the Qualtrics and Rainloop-based testing platforms, participants were directed to an additional Qualtrics-based interactive module with four additional communications: two letters (one fake, one real) and two text messages (one fake, one real). These non-email messages also allow us to assess how well the training generalized to other types of potentially fraudulent communication.

3.3 Research Questions and Extensions

The study, as originally proposed, had a single hypothesis: *participants who are exposed to a weakened version of an actual email scam will be more likely in a subsequent test to correctly distinguish scam from non-scam emails relative to a control group.* This hypothesis was to be tested in a randomized control trial with three experimental arms, in which the sample size for the study was 2,000 participants, with an outcome variable as the percent of correct answers, and a minimum detectible effect of 9% (power = .9; alpha = 0.05; two-sample test of proportions; assumed baseline for control of 50% correct).

However, in the initial trials of the intervention, it became rapidly clear that the effect was far more than that minimum detectible effect. It also became clear that the platform developed to test that hypothesis could be used more broadly, at no additional cost. Thus, in the early months of the study, additional research questions were added. It is vital for the credibility of the research community that researchers separate prior hypotheses from ex post and contemporaneous areas of

interest. In the result below, the measurement of the inoculation's effectiveness on the number of correctly identified messages is a correctly specified experimental analysis. All other analyses should be considered exploratory in nature – even when conducted via randomized control trials.

In total, this study addresses the following research questions:

R1: Does an inoculation using weakened scams increase the likelihood that individuals will correctly identify subsequent scams?

R2: Does the impact of the training, if any, generalize across communication mediums (email to SMS or email to letter)?

R3: Does the impact of the training, if any, generalize across who is being impersonated?

R4: How quickly does the effectiveness, if any, of the training dissipate over time?

R5: Does the impact of the training, if any, generalize across user interfaces – or is it specific to the look and feel of the training scenario?

R6: What personal characteristics predict fraud susceptibility?

3.4 Data Collection

Participants were sourced from a commercial provider of online survey participants, Prolific. Prolific is a higher-quality alternative to the commonly used Mechanical Turk service and allows for the creation of nationally representative samples. Panel participants were paid to participate and compensated based on the study's duration; the median compensation was $9/hr. More information about the Prolific service, and how it generates a nationally representative sample of US residents can be found in Appendix C. Multiple experiments were conducted across separate random samples, for a total sample size of 4,164 participants;[5] the size and characteristics of each random sample is noted in the results section below. For completeness, detailed information about all samples drawn from the population is discussed in Appendix C.

[5] As noted above, the study expanded from an initial target of one experiment with 2,000 participants to include multiple experiments with 4,164 participants. Each subsequent sample excluded prior participants. This includes an unsuccessful attempt to use another panel provider, Dynata, whose data had to be discarded because of the low quality of the responses, as explained in Footnote 8.

For each sample, participant interactions with the training and testing were tracked and stored in an anonymous database: a Qualtrics backend for the relevant portions and a DynamoDB backend for the Rainloop interactions. Python code is provided on the project's GitHub site that processes the various data sources, combines them into a single data file, and runs the analyses.

4. Results

The study provided rich opportunities for understanding not only the experiment's stated goal – helping individuals detect fraud – but also how various factors such as the delay period, medium of communication, and testing platform affected participant outcomes.

4.1 Research Question 1: Does an inoculation using weakened scams increase the likelihood that individuals will correctly identify subsequent scams?

The analysis starts with a nationally representative sample of 1,065 residents of the United States, in which the training and test were both conducted in Qualtrics. All delay periods are included, and their answers on all twelve communications are considered. 712 out of the 1,065 participants completed both rounds of the study and passed the attention checks.

In this experiment, the interactive training shows a statistically significant increase over the control in the number of correct answers offered by participants ($p<0.001$). None of the other arms show statistically significant results relative to the control. The average treatment effect of the interactive training is 0.82 additional questions (out of 12) answered correctly on the study, resulting in 8.1 correct answers in the control and 8.9 correct answers with the training, as shown in Table 1.

Table 1: Results for Research Question 1, Does the training help people correctly identify scams? N=712, Qualtrics-based test and training. Values shown in the format: average (standard deviation).

Outcome	Control	Written Tips	Existing Communication	Interactive Training
Number Correctly Labeled	8.11 (1.82)	8.10 (1.63)	7.92 (1.68)	8.93 (1.68)***
Number of Fake Messages Labeled Real	2.24 (1.49)	2.14 (1.39)	2.41 (1.49)	1.81 (1.32)**
Number of Real Messages Labeled Fake	1.65 (1.18)	1.76 (1.09)	1.68 (1.18)	1.26 (1.02)**

*Symbols: *** p<0.001; ** p<0.01; * p<0.05. Two-sided t-test on the number of communications correctly identified, relative to the control condition. The initial experimental design (number of correctly labeled communications) shared a common control and did not require a Family-wise Type 1 error correction; however, out of an abundance of caution and since the additional exploratory research questions have been added, the authors also applied a Holm (Bonferroni) adjustment to the p values and found that the results are still significant (alpha of 5%).*

It is not only important that individuals can correctly identify fraud, it is also essential that individuals can identify and trust *real* communications. Here again, the results are encouraging. The average treatment effect of 0.82 correct answers comes from both learning to better identify

fakes and to better identify real messages. Specifically, the interactive training increases the correctly identified fakes by 0.43 relative to the control ($p<0.01$); no other arms are statistically significant. In addition, the interactive training increases the correctly identified real messages by 0.39 relative to the control ($p<0.01$); no other arms are statistically significant. The resulting values are also provided in Table 1.

In two further checks for robustness, the results hold as well. First, the interactive training shows a statistically significant improvement relative to the two other, non-control arms. Second, the interactive training is significant when the same analysis is conducted as a multivariate OLS regression with additional control variables (employment status, generalized trust, prior experience with fraud, level of education, marital status, age, gender, time delay between first and second rounds) as a check against incomplete randomization.

While the statistical significance of the results is straightforward to interpret, the practical significance is more difficult: the magnitude of the training's effect is largely determined by the mix of real and fake messages and how difficult they are to identify correctly. The message mix used in the study is not representative of a real-world scenario; instead, the mix was intentionally chosen to create a diversity of messages and levels of difficulty. One can see this by looking at the accuracy rate for each message, given in Table 2:

Table 2: Impact of Interactive Training, by Message. N=712, Qualtrics-based test and training.

Message	% Correct: Control	% Correct: Interactive Training	Difference (Percentage Points)	% Change	Significance
Your Urgent Support Is Needed	42%	63%	21%	50.0%	p:0.000; ***
Need a replacement Social Security Card?	64%	85%	21%	32.8%	p:0.000; ***
Important Information About Your Online Account	79%	91%	12%	15.2%	p:0.003; **
The Social Security Administration is contacting a few people	47%	59%	12%	25.5%	p:0.028; *
Payment declined: Update your information so we can ship your order	61%	70%	9%	14.8%	p:0.058
Opt Out of Receiving Mailed Notices	43%	49%	6%	14.0%	p:0.307
Annual Reminder to Review Your Social Security Statement	68%	74%	6%	8.8%	p:0.192
Disability Alert	94%	97%	3%	3.2%	p:0.235
Delivery Update	93%	94%	1%	1.1%	p:0.678
Notice of Intent to Levy Social Security Benefits	82%	81%	-1%	-1.2%	p:0.741
The Application Process Is Open For Disability Benefits	86%	85%	-1%	-1.2%	p:0.884
Convalescent Plasma is needed	52%	46%	-6%	-11.5%	p:0.292
Average	68%	75%	7%	10.2%	

*Symbols: *** p<0.001; ** p<0.01; * p<0.05. Two-sided test of proportions (z test) on the number of communications correctly identified, relative to the control condition.*

There is a wide variation in both the baseline rates of correct answers in the control condition (from 42% to 92%). This represents the level of difficulty of the message. There is also a wide variation in the impact of the training, from 21 percentage points (statistically significant) to a -6% percentage point *decrease* (not statistically significant, and likely noise). An analysis of the average impact alone does not capture that variation adequately. Again, this variation is intentional and useful for the study since it allows us to better understand the *types* of messages for which the training is effective; we analyze those questions in subsequent sections of the report. In terms of practical significance, we can say that the training, when it has a statistically significant impact, boosts correct responses from 12 to 21 percentage points (a 15% to 50% difference). Across this *particular mix of messages*, there is a statistically significant impact of 7 percentage points or a 10% increase in correct answers.

4.2 Research Question 2: Does the impact of the training, if any, generalize across communication mediums (email to SMS, or email to Letter)?

The interactive training focused on email communications, with five emails. It also included one SSA letter to broaden the participants' awareness. The testing process included eight emails (primary training), two letters (partial training), and two SMSes (no training). Table 3 below shows the experimental results by medium of communication.

Table 3: Results for Research Question 2, Does the impact of the training generalize across communication mediums?
N=712, Qualtrics-based test and training. Values shown in the format: average (standard deviation).

Outcome	Control	Written Tips	Existing Communication	Interactive Training
Number Emails Correctly Labeled	5.36 (1.51)	5.46 (1.49)	5.20 (1.47)	6.10 (1.40)***
Number SMSes Correct	1.46 (0.53)	1.40 (0.52)	1.41 (0.55)	1.43 (0.52)
Number Letters Correct	1.29 (0.59)	1.24 (0.58)	1.31 (0.55)	1.40 (0.61)

*Symbols: *** p<0.001; ** p<0.01; * p<0.05. Two-sided t-test, relative to the control condition. With Holm (Bonferroni) adjustment, the results are still significant (alpha of 5%).*

The primary effect of the interactive training is on the emails: with a boost of 0.74 messages correctly identified (p<0.001). For letters, there is suggestive evidence (p=.1), that perhaps would

be statistically meaningful if more letters had been included in the training and test processes. For SMSes, there was no effect.

Thus, it appears that the training works for the communication medium people are trained on, but not others. That should not be surprising, given some of the most important techniques participants learned to identify scams are specific to the email: hovering over URLs and looking at email's headers, for example. These techniques are irrelevant for letters and not supported by most SMS applications.

In case of incomplete randomization, the research team also ran the analysis as a regression with additional controls for gender, age (quadratic), log income, years of education, employment status, prior loss of money to fraud, and generalized trust score. The average treatment effect of the experiment is slightly larger than presented above (0.75 additional messages correct), and similarly, it is statistically significant.

4.3 Research Question 3: Does the impact of the training, if any, generalize across who is being impersonated?

The training process included only messages from (or appearing to come from) the Social Security Administration. The test included messages both from the Social Security Administration and from other parties: the Red Cross and Amazon. Amazon impostor scams have increased in frequency dramatically in 2021, with 100 to 150 million calls per month observed in February, March, and April of 2021 (YouMail 2021). We can see in Table 4 that the impact of the training does generalize to non-SSA related messages.

Table 4: Results for Research Question 3, Does the impact of the training generalize across the content of the imposter scam?
N=712, Qualtrics-based test and training. Values shown in the format: average (standard deviation).

Outcome	Control	Written Tips	Existing Communication	Interactive Training
Number Correctly Labeled: SSA	3.99 (1.09)	4.04 (1.04)	3.82 (1.14)	4.44 (1.09)***
Number Correctly Labeled: Non-SSA	4.12 (1.25)	4.06 (1.17)	4.10 (1.07)	4.49 (1.02)**
Number Emails Correctly Labeled: SSA	1.76 (0.83)	1.83 (0.91)	1.60 (0.93)	2.08 (0.82)***
Number Emails Correctly Labeled: Non-SSA	3.61 (1.12)	3.64 (1.05)	3.60 (1.01)	4.03 (0.95)***

*Symbols: *** p<0.001; ** p<0.01; * p<0.05. Two-sided t-test, relative to the control condition. With Holm (Bonferroni) adjustment, the results are still significant (alpha of 5%).*

4.4 Research Question 4: Does the impact of the training, if any, generalize across user interfaces?

The results discussed thus far used Qualtrics for both the training and testing process. The next study uses Qualtrics for the training, but the Rainloop platform (discussed above) for the test. This study entails a 1,500 person nationally representative survey, with participants randomly assigned into two groups[6]: a two-week delay period and a four-week delay period. Here, the results are drawn from the first wave, a two-week delay with 533 people who completed both the test and training parts of the study and passed the attention checks.

This study differs from the prior work in two important ways. First, in using Rainloop, we can analyze the generalizability of the training to a very different visual design and user interface: one intentionally designed to mimic a standard email reader. Second, in Rainloop's preview pane, the senders are identified by their names only and not by their email domain. This is the standard practice in email readers such as Gmail but was not the case in either the training or Qualtrics-based test. In testing Rainloop, the authors discovered that it was trivially easy for participants to spot the fakes when the email domain was displayed in the preview pane.[7,8] Thus, the team switched to a more realistic, name-only display of the sender.

---

[6] The researchers started a 2,000-respondent sample with Dynata using the Rainloop platform before realizing that Dynata had significant data quality issues. Dynata is a global market research company originally founded in 1999 and previously known as Research Now Survey Sampling International (SSI). It is a long-standing provider of market research tools and survey participants, including for academic studies. Whereas the prior Prolific study had a 68 percent second-round completion rate (68 percent of people who started the first study completed through the end of the second study with apparently valid data), the study with Dynata had a 15 percent second-round completion rate. The main reason, however, is that only 24 percent of those who started the second round provided valid data: most of them simply did not open the test emails, as instructed. For the few who remained, their data appears to be highly noisy. We had to discard the data from Dynata altogether.

[7] Specifically, when the domains for each message are displayed altogether, it is clear that some are from a .gov domain while others are not. The Qualtrics-based test displayed messages on separate pages, making it more difficult to notice such inconsistencies. Again, in a live email environment the domains would not be displayed together; hence, the design change made it more realistic.

[8] The team ran a non-representative sample of 438 participants from Prolific (Sample '5P', as described in Appendix A) through the original Rainloop design with sender domains with one key finding: both the interactive training and the non-interactive written tips had a statistically significant effect on currently identifying emails (with no effect on Letters or SMSs). There was no statistically significant difference between the interactive training and written tips.

The results in Rainloop are quite similar as for the previous Qualtrics study and are provided in Table 5. The practically and statistically significant impact of interactive training remains with a different user interface. This suggests that the lessons are not specific to the training environment and provide broader protection. As before, we find that the impact occurs especially with emails: participants learn how to identify fraudulent and real emails correctly, but not SMSes or letters.

Table 5: Results for Research Question 4, Does the impact of the training, if any, generalize across user interfaces? N=533 completes, Qualtrics-based test and Rainloop-based training. Values shown in the format: average (standard deviation).

Outcome	Control	Written Tips	Existing Communication	Interactive Training
Number of Communications Correctly Labeled	7.38 (1.90)	7.48 (1.90)	7.49 (1.85)	8.09 (2.01)**
Number of Fake Communications Labeled Real	3.06 (1.73)	2.85 (1.71)	2.86 (1.69)	2.50 (1.81)**
Number of Real Communications Labeled Fake	1.45 (1.14)	1.41 (1.03)	1.44 (1.05)	1.27 (1.02)
Number Emails Correctly Labeled	4.68 (1.48)	4.91 (1.65)	4.88 (1.43)	5.41 (1.79)***
Number SMSes Labeled	1.50 (0.59)	1.42 (0.64)	1.34 (0.68)*	1.40 (0.57)
Number Letters Labeled	1.20 (0.69)	1.16 (0.68)	1.28 (0.75)	1.27 (0.63)

*Symbols: *** p<0.001; ** p<0.01; * p<0.05. Two-sided t-test, relative to the control condition.*

The Rainloop platform also allows us to track individual actions in detail. For example, in the interactive training, participants opened and reviewed twice as many email headers as in the control group (2.36 versus 1.12; p<.001). One of the key lessons taught during the interactive training is to review the email headers.

4.5 Research Question 5: How quickly does the effectiveness, if any, of the training dissipate over time?

Within the Qualtrics-based study, the authors included two different delay periods between the test and control. Participants were recruited for the training process and then were randomly invited to participate in the testing process either three days from the start of the training recruitment or ten days after. However, the nature of online recruitment makes this process imprecise – since once a study is posted for participants, participants can read that notice and decide to complete the study whenever they like. The actual, realized delay period between the training and the test ranged between 1.3 days to 12.8 days.[9]

[9] Even though the test was first posted three days after the training, a small number of people took the training late then took the test shortly afterward.

This participant choice complicates our analysis on the test delay because it introduces a confounder into an otherwise clean experiment: there is likely an unseen omitted variable (personal characteristics) that both cause the person to select a certain delay period and respond differentially to the training. Only at the aggregate level (across all delay periods) does the randomization process holds and one can trust the results given above.

Given those caveats, one can estimate the effect of the training test delay by grouping people into two waves: those who were (randomly selected to be) *invited* to participate in the test three days after the start of the training and those who were invited to participate ten days after.

In this initial analysis, the effect of interactive training diminishes with time. In the first group, with a median actual delay period between of 3.5 days between the training and the test, the training increased the number of correct answers by 1.1 ($p<0.001$). In the second group, with a median realized delay period of 11.5 days, the training increased the number of correct answers by 0.56 ($p=0.02$). Both are practically and statistically significant, and no other arms showed a statistically significant effect, but the effect wanes by the second period.

Another way of analyzing the effect of time is to look at the interaction effect between the experimental arm and the Wave, along with additional control for the time delay between the treatment and test. The OLS regression uses the following form:

$$\begin{aligned}\textit{Num Emails Correct} = {} & \beta_0 + \beta_1\, \textit{ReceivedTips} + \beta_2\, \textit{ReceivedExistingCommunication} + \\ & \beta_3\, \textit{ReceivedInteractiveTraining} + \beta_4\, \textit{Wave} + \beta_5\, \textit{ReceivedTips*Wave} + \beta_6\, \textit{ReceivedExistingCommunication*Wave} + \\ & \beta_7\, \textit{ReceivedInteractiveTraining*Wave} + \beta_8\, \textit{DaysFromTrainingToTest} + \varepsilon\end{aligned}$$

The results show that none of the coefficients are statistically significant, except for the previously discussed effect of the training itself (here: $\beta_3 = .99$; $p<0.001$). There are hints of a negative effect on the Wave (the more time elapsed between the invitations, correct answers decrease) and a positive effect on days between training and test (the longer people took to respond to the invitation, the more correct answers). Both are not statistically significant, however.

In a second test of the effect of time delays, we extended the Rainloop test mentioned above to include an additional 692 (randomly selected) participants were invited to take the test module four weeks after their training. Only 434 completed the training module and passed the attention check, which serves as a warning sign that the results may be skewed towards participants who are frequent users of Prolific and still active on the site four weeks later. That said, the results are in line with the earlier time-delay test and show an additional decrease in the effect. Specifically,

the results for the interactive training show an average improvement over all other arms, but the results are no longer statistically significant. Further work is certainly needed, but the results thus far indicate what one would expect: the impact of the training appears to decay over time.

4.6 Research Question 6: What personal characteristics predict fraud susceptibility?

Within a respondent's incorrect answers, we can separate two effects: the effects of fraud susceptibility (labeling a fake message as real) and undue distrust (labeling real messages as fake).

As noted above, a challenge in the existing literature on fraud susceptibility is that it is observational: it relies on the self-selected responses of individuals who were (selectively) targeted by fraudsters, (selectively) tricked by the fraud attempt, (selectively) who knew they were targeted by fraud, and (selectively) reported the fraud attempt. Only the second stage – being tricked by the fraud – is truly fraud susceptibility; the others are practically important but distinct elements of the fraud ecosystem.

This study allows us to look at the problem from a very different angle; a nationally representative sample, shows who is tricked by the fraud attempt and removes the other confounding factors. To do so, we run a series of analyses on the number of fraudulent messages that participants labeled as real. To be clear, this is an exploratory analysis and not a theoretically driven, pre-registered test of hypotheses. The existing literature indicated that age and generalized trust might be factors, but we (the authors of this report) did not have a sufficient basis to pre-register a set of hypotheses. Thus, this analysis can point us to additional research but should not be taken as definitive. Now that the caveats have been properly covered, Table 6 provides the results.

Table 6: Exploratory Regression of Personal Characteristics on Fraud Susceptibility
N=712, Qualtrics-based test and training. Values shown in the format: average (standard deviation).

Variable	Coefficient	Standard Error	P>|t|
Intercept	5.31	0.75	0.000
Received Tips	-0.05	0.15	0.764
Received Existing Communication	0.15	0.16	0.327
Received Interactive Training	-0.41	0.15	0.008
Employment Status = Retired	0.08	0.18	0.655
Employment Status = Unemployed	0.10	0.13	0.455
Lost Money To Fraud	-0.16	0.35	0.655
Days From Training To Test	-0.01	0.02	0.683
Generalized Trust Score	-0.06	0.04	0.146
Log(Income)	0.08	0.08	0.267

Years of Education	-0.06	0.03	0.024
Married [Married=1; Not = 0]	0.10	0.13	0.441
Age	-0.10	0.03	0.000
Age^2	0.00	0.00	0.001
Gender [Female=1; Male=0]	0.01	0.11	0.899
Time Used to Complete the Test (Quantile)	-0.05	0.04	0.273

Here, we find that fraud susceptibility decreases with age, but with a curve: the peak of the curve (the least likely age to fall prey to fraud) is 53 years old, with a decrease of -2.7 fraudulent messages thought to be real, all else being equal. Before and after that age, the effect shrinks. People at 20 and age 89 are equally likely, all else constant, to fall prey to fraud. In addition, the effect of the interactive training remains, as expected. The only other new statistically significant finding we see is that susceptibility to fraud decreases with the person's years of education. Someone who has completed a four-year college is expected to fall prey to 1.01 fewer fraudulent messages, all else being equal.

Next, let us look at *undue distrust:* who tends to label real messages as fraudulent? For this, we run the same regression but change the dependent variable to the number of real messages thought to be fraudulent. The results can be found in Table 7.

Table 7: Exploratory Regression of Personal Characteristics on Undue Distrust
N=712, Qualtrics-based test and training. Values shown in the format: average (standard deviation).

Variable	Coefficient	Standard Error	P>|t|
Intercept	2.21	0.58	0.000
Received Tips	0.09	0.12	0.446
Received Existing Communication	0.02	0.12	0.859
Received Interactive Training	-0.38	0.12	0.002
Employment Status = Retired	-0.06	0.14	0.692
Employment Status = Unemployed	-0.16	0.10	0.123
Lost Money To Fraud	-0.13	0.28	0.637
Days From Training To Test	-0.01	0.01	0.317
Generalized Trust Score	-0.08	0.03	0.021
Log(Income)	-0.12	0.06	0.045
Years of Education	-0.02	0.02	0.452
Married [Married=1; Not = 0]	0.34	0.10	0.001
Age	0.01	0.02	0.529
Age^2	0.00	0.00	0.829
Gender [Female=1; Male=0]	-0.16	0.09	0.062
Time Used to Complete the Test (Quantile)	-0.04	0.03	0.219

Here, we find that the 3-question generalized trust score is important: the higher one scores on generalized trust, the less likely they are to show undue, inaccurate distrust, all else being equal.

Similarly, we find that women, all else being equal, are less likely to show undue distrust; men are more distrusting on this exercise. The higher the person's income, all else being equal,

the less undue distrust. Finally, people who are married show greater distrust. These results are fascinating and provide fodder for future analyses; remember, however, that this is an exploratory analysis, so we should be careful before drawing broad conclusions about society.

5. Conclusion

Fraud schemes and imposters that impersonate the Social Security Administration are a serious problem affecting millions of people in the United States. This study uses a set of randomized control trials to test the effectiveness of new and existing approaches to train people to spot fraud. An informational message currently used by the Social Security Administration to warn people about fraud shows no effect, nor does a set of written instructions on how to spot fraud.

The training that does show promise is an interactive training, in which individuals directly experience fraudulent communications and are *inoculated* against them. The interactive training provides a statistically significant improvement in the participant's ability to correctly identify communications as fraudulent or not. The average effect based on the particular mix of messages used in the study is seven percentage points (a 10% increase). In comparison, the effect for individual messages ranges from non-significant to 21 percentage points (up to a 50% increase from the control case).

The effect comes from teaching to people to detect fraudulent messages and teaching them to trust real messages. The training focuses on practical techniques to use with emails, and its impact is specific to that medium. There is no statistically significant effect for SMS or letters, which is not surprising since the techniques taught to identify fraudulent emails (looking at email headers and links) are not relevant for those mediums. The effect of the training generalizes to other user interfaces. A different email platform and user interface were used in one of the tests, with similar results. The effect of the training also can generalize to non-SSA messages; participants were able to correctly label fraudulent and real messages from the Amazon and Red Cross, for example. The duration of the training is limited though – decreasing after two weeks and losing statistical significance after four weeks.

In an exploratory analysis, the authors find that fraud susceptibility decreases with age but with a curve: the peak of the curve (the least likely age to fall prey to fraud) is 53 years old. Susceptibility to fraud also decreases with the person's years of education. Another exploratory analysis looks at who is subject to *undue distrust*, or when people tend to label real messages as fraudulent. Women, people with higher income, and those with higher scores on a generalized trust measure are all less prone to undue, inaccurate distrust.

This study suggests that simple, online training programs can be designed to help reduce imposter scams. These interventions do not necessarily have to be expensive to be effective. For example, the training process used in this study took participants a median of 4.6 minutes to complete and was developed at a very low cost using a few hours of a professional designer's time. The primary cost to deploy this intervention at scale would be the cost of marketing it to individuals. If this training were added to existing interactions with the public, such as when people register on SocialSecurity.gov, the marketing cost would be minimal.

This study also points to many questions that still need to be explored in terms of optimal design and structuring interventions for multiple mediums through which fraudsters might seek to contact consumers. This research can be accomplished using this, or a similar, experimental research platform to provide rigorous results that have been difficult to obtain in the past using victim-reported fraud data.[10]

[10] To empower other researchers, all elements of the study are now Open Source: the two platforms developed to power the research, in Qualtrics and in Rainloop, the data resulting from the use of these platforms, and the code used to analyze the resulting data. Each of these are all freely available to the research community at the author's GitHub site, with accompanying documentation, at https://github.com/sawendel/ssascams. The authors look forward to an active conversation with other members of the research community so that we might together better tackle this significant threat to people's well-being and the proper functioning and success of government bodies such as the Social Security Administration.

6. References

6.1 Research

American Association of Retired Persons. 2019. “Social Security Scams.” AARP. Accessed April 18, 2020. http://www.aarp.org/money/scams-fraud/info-2019/social-security.html.

Anandpara, Vivek, Andrew Dingman, Markus Jakobsson, Debin Liu, and Heather Roinestad. 2007. “Phishing IQ Tests Measure Fear, Not Ability.” In *Financial Cryptography and Data Security*, edited by Sven Dietrich and Rachna Dhamija, 362–66. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-77366-5_33.

Banas, John A., and Stephen A. Rains. 2010. “A Meta-Analysis of Research on Inoculation Theory.” *Communication Monographs* 77 (3): 281–311. https://doi.org/10.1080/03637751003758193.

Burnes, David, Charles R. Henderson, Christine Sheppard, Rebecca Zhao, Karl Pillemer, and Mark S. Lachs. 2017. “Prevalence of Financial Fraud and Scams Among Older Adults in the United States: A Systematic Review and Meta-Analysis.” *American Journal of Public Health* 107 (8): e13–21. https://doi.org/10.2105/AJPH.2017.303821.

Chen, Hongliang, Christopher E. Beaudoin, and Traci Hong. 2017. “Securing Online Privacy: An Empirical Test on Internet Scam Victimization, Online Privacy Concerns, and Privacy Protection Behaviors.” *Computers in Human Behavior* 70 (May): 291–302. https://doi.org/10.1016/j.chb.2017.01.003.

Compton, Josh. 2013. “Inoculation Theory.” In *The SAGE Handbook of Persuasion: Developments in Theory and Practice*, edited by James Price Dillard and Lijiang Shen, 220–37. Los Angeles: SAGE Publications

Coppock, Alexander, and Oliver A. McClellan. 2019. "Validating the Demographic, Political, Psychological, and Experimental Results Obtained from a New Source of Online Survey Respondents." *Research & Politics* 6 (1). https://doi.org/10.1177/2053168018822174.

Federal Bureau of Investigation: Internet Crime Compliant Center. 2021. "Internet Crime Report 2020." Accessed 16 September 2021. https://www.ic3.gov/Media/PDF/AnnualReport/2020_IC3Report.pdf.

Federal Trade Commission. 2021. "Consumer Sentinel Network Data Book 2020." *Washington, DC*. Accessed 16 September 2021. https://www.ftc.gov/reports/consumer-sentinel-network-data-book-2020.

FINRA Investor Education Foundation. 2013. "Financial Fraud and Fraud Susceptibility in the United States. Research Report from a 2012 National Survey." Applied Research and Consulting New York, NY. Accessed 16 September 2021. https://www.saveandinvest.org/sites/saveandinvest/files/Financial-Fraud-And-Fraud-Susceptibility-In-The-United-States.pdf.

Fletcher, Emma. 2019. "Growing Wave of Social Security Imposters Overtakes IRS Scam." Federal Trade Commission. Last Updated 12 April 2019. https://www.ftc.gov/news-events/blogs/data-spotlight/2019/04/growing-wave-social-security-imposters-overtakes-irs-scam.

Holtfreter, Kristy, Michael D. Reisig, and Travis C. Pratt. 2008. "Low Self-Control, Routine Activities, and Fraud Victimization*." *Criminology* 46 (1): 189–220. https://doi.org/10.1111/j.1745-9125.2008.00101.x.

KLEW. 2020. "Scam Alert: Phishing Email Appears to Come from Social Security Administration." Last Updated 6 May 2020. https://klewtv.com/news/local/scam-alert-phishing-email-appears-to-come-from-social-security-administration.

Leach, Jennifer. 2018. “Fake Calls about Your SSN.” Consumer Information. Last Updated 12 December 2018. https://www.consumer.ftc.gov/blog/2018/12/fake-calls-about-your-ssn.

McGuire, William J. 1961. “The Effectiveness of Supportive and Refutational Defenses in Immunizing and Restoring Beliefs Against Persuasion.” *Sociometry* 24 (2): 184–97. https://doi.org/10.2307/2786067.

McGuire, William J, CC Haaland, and WO Kaelber. 1964. “Inducing Resistance to Persuasion. Some Contemporary Approaches.” In *Advances in Experimental Social Psychology*, edited by Leonard Berkowitz, 1:191–229.

Muscat, Glenn, Marianne James, and Adam Graycar. 2002. “Older People and Consumer Fraud.” 220. Trends & Issues in Crime and Criminal Justice. Canberra: Australian Institute of Criminology. http://www.aic.gov.au/publications/tandi/ti220.pdf.

Roozenbeek, Jon, and Sander van der Linden. 2019. “Fake News Game Confers Psychological Resistance against Online Misinformation.” *Palgrave Communications* 5 (1): 1–10. https://doi.org/10.1057/s41599-019-0279-9.

Saleh, Nabil F., Jon Roozenbeek, Fadi A. Makki, William P. Mcclanahan, and Sander Van Der Linden. 2021. “Active Inoculation Boosts Attitudinal Resistance against Extremist Persuasion Techniques: A Novel Approach towards the Prevention of Violent Extremism.” *Behavioural Public Policy*, 1–24. https://doi.org/10.1017/bpp.2020.60.

Scheithe, Erin. 2020. “Five Ways to Recognize a Social Security Scam.” Consumer Financial Protection Bureau. Last Updated February 18, 2020. https://www.consumerfinance.gov/about-us/blog/five-ways-to-recognize-social-security-scam/.

SimplyWise. 2021. “SimplyWise Retirement Confidence Index.” *SimplyWise* (blog). Last Updated 18 January 2021. https://www.simplywise.com/blog/retirement-confidence-index/.

Skiba, Katherine. 2021. “Social Security Impostor Complaints Break Record in 2020.” AARP. Last Updated 4 March 2021. https://www.aarp.org/money/scams-fraud/info-2021/record-high-social-security-impostor-complaints.html.

Titus, Richard M., Fred Heinzelmann, and John M. Boyle. 1995. “Victimization of Persons by Fraud.” *Crime & Delinquency* 41 (1): 54–72. https://doi.org/10.1177/0011128795041001004.

Van Wyk, Judy, and Michael L. Benson. 1997. “Fraud Victimization: Risky Business or Just Bad Luck?” *American Journal of Criminal Justice* 21 (2): 163–79. https://doi.org/10.1007/BF02887448.

Waggoner, John. 2020. “Social Security Scammers Turn to Email.” AARP. Last Updated 9 January 2020. https://www.aarp.org/money/scams-fraud/info-2020/social-security-email.html.

Wendel, Stephen. 2020. *Designing for Behavior Change: Applying Psychology and Behavioral Economics*. 2nd Edition. Sebastopol, California: O’Reilly Media.

Whitty, Monica T. 2019. “Predicting Susceptibility to Cyber-Fraud Victimhood.” *Journal of Financial Crime* 26 (1): 277–92. https://doi.org/10.1108/JFC-10-2017-0095.

YouMail. 2021. “Amazon Imposter Robocalls Reaching 150 Million Per Month.” Last Updated 13 May 2021. https://www.prnewswire.com/news-releases/amazon-imposter-robocalls-reaching-150-million-per-month-301290910.html.

6.2 Software

- GoPhish: Open Source Phishing Framework. https://getgophish.com/
- RainLoop: https://www.rainloop.net/
- Qualtrics: https://www.qualtrics.com/
- DynamoDB, part of Amazon Web Services: https://aws.amazon.com/dynamodb/

Appendix A:
Supplementary Detail on the Experiment Design

The following information be found online at https://github.com/sawendel/ssascams:

1. A guide to the data, for analysis by other researchers
2. An overview of the research platform used in the study, which is open source and available for other researchers to use.

Table A-8 provides an overview of each message from the experiment; the full contents of each message can be found on the Github Site under "html\v2akureytemplate\[test|training]".

Table A-8: Communications Used in the Experiment

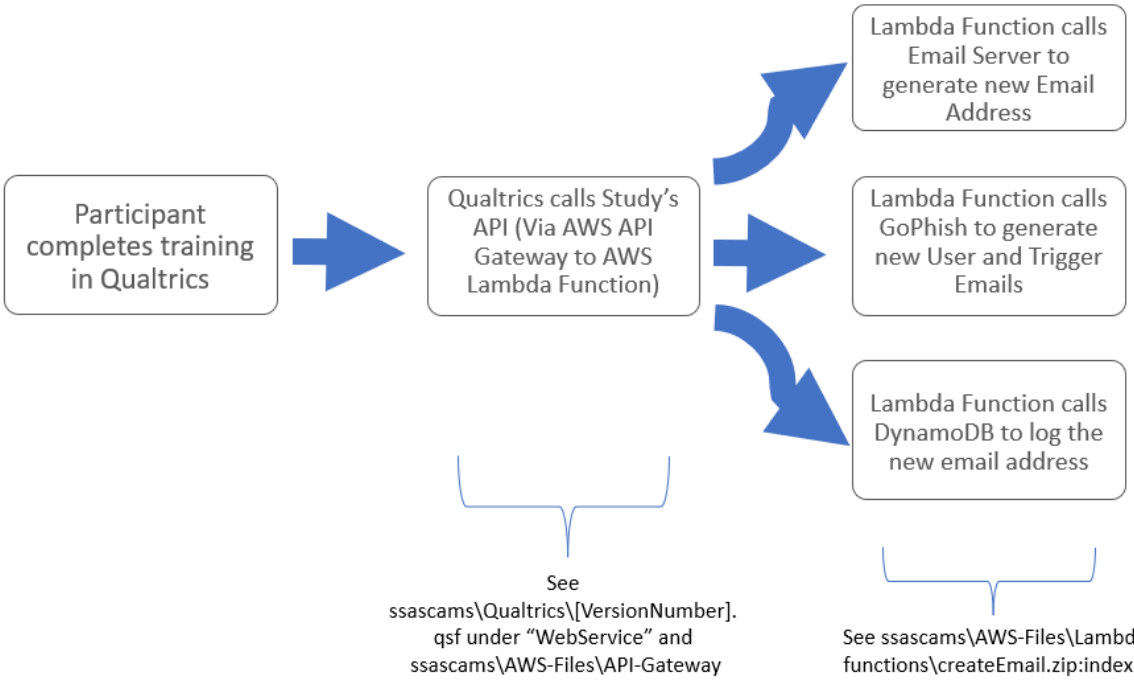
Medium	Real or Fake?	Subject Line	From Name	From Address	Reply-To
Email	Real	Important Information About Your Online Account	NO-REPLY@ssa.gov	NO-REPLY@ssa.gov	NO-REPLY@ssa.gov
Email	Fake	Payment declined: Update your information so we can ship your order	Amazon.com	payments-update@gmail.com	payments-update@gmail.com
Email	Real	Delivery Update	Amazon.com	no-reply@amazon.com	order-update@amazon.com
Email	Fake	Your Urgent Support Is Needed	Red Cross Covid Campaign	covidcampaign@redcross.biz	covidcampaign@redcross.biz
Email	Fake	The Application Process Is Open For Disability Benefits	Application Assistance Program	applicationassistance@disabilitybenefits.com	disabilityhelp@gmail.com
Email	Fake	Opt Out of Receiving Mailed Notices	NO-REPLY@ssa.security	notices@ssa.security	notices@ssa.security
Email	Real	Need a replacement Social Security Card?	NO-REPLY@ssa.gov	NO-REPLY@ssa.gov	NO-REPLY@ssa.gov
Email	Fake	Annual Reminder to Review Your Social Security Statement	NO-REPLY@social.security.org	NO-REPLY@social.security.org	NO-REPLY@social.security.org
Letter	Real	The Social Security Administration is contacting a few people…	Social Security Administration; Office of Quality Review	NA	NA
SMS	Fake	Disability Alert	949-409-0220	NA	NA
Letter	Fake	Notice of Intent to Levy Social Security Benefits	Benefits Suspensions Unit, Arlington County	NA	NA
SMS	Real	RED CROSS: Convalescent Plasma is needed	909-99	NA	NA

Appendix B: About the Technical Platform

Overview of the Environment

The participant interactions occur in two phases: training and test. The backend for the training process is depicted in Figure B-7.

Figure B-7: Backend Flow for Training Process



The testing process is straightforward. First, the user goes to the Qualtrics survey for the test. In the survey, the user activates a new window which shows the Rainloop interface, with their username and password automatically filled in. Within Rainloop, all actions are logged to the DynamoDB. After the person has finished interacting with the emails in Rainloop, they close that window and complete the survey in Qualtrics.

Notes about Rainloop

For this project, the research team modified the source code of Rainloop to create a safe testing environment where participants could interact with it as they would with any other email program, but all external communication and consequences were removed. Participants could

open and close the messages, view headers, click reply, delete messages, or label them as spam. However, these actions were each limited in their effect; for example, they could compose a message, but not actually send the email outside of the environment.[11] Links were disabled so that the user could click on them but could not actually be directed to external websites.

Notes about GoPhish

The emails were designed and implemented in GoPhish[12], an open-source phishing security tool used by companies to test the susceptibility of their employees to phishing attacks. GoPhish provides an administrative interface for researchers and security professionals to setup realistic phishing campaigns, target sets of users, and track the results of those campaigns. In this case, the researchers modified GoPhish's code to work seamlessly with the Qualtrics training module and Rainloop testing module. An external design and engineering firm, Akurey, customized the tools and provided graphic design for the scam emails.

[11] The text of their messages was not logged for privacy reasons. Only the fact that they clicked "reply" was logged, using their otherwise anonymous user id (the ID provided by Prolific or Dynata).

[12] Available at https://getgophish.com/.

Appendix C: About Prolific and the Prolific Samples

This study uses Prolific to collect its sample of US residents. Prolific is a more recent and upcoming competitor to the commonly used Mechanical Turk service. It was founded in 2014 by Ph.D. students at Oxford and is based in the UK. They cater to the academic market, with a particular focus on supporting "ethical and trustworthy research".[13] Overall, the authors of this study have found that the panel services Prolific provides are of significantly higher-quality and reliability than Amazon's Mechanical Turk service.

A number of studies have been published that analyze the characteristics of these panel providers, including the demographics and behavior of their respondents. For example, Coppock and McClellan (2019) analyze the service from Lucid, another panel provider in the field, and find remarkable congruence with US demographics and with psychological responses from published social science research. A recent analysis of Mechanical Turk, Qualtrics Panels, Dynata, and Prolific found that "only Prolific provided high data quality on all measures" and "MTurk showed alarmingly low data quality even with data quality filters" (Peer et al. 2021). It should be noted that some of the authors on the paperwork at Prolific, however, as this sort of conflict could have influenced the results.

What Prolific and similar providers offer is a nationally representative, quota-based sample at a reasonable price.[14] In a quota-based sample, potential participants are selected from a broader (non-representative) pool of people until specific quotas are met by age, gender, and other demographics so that the resulting sample matches those characteristics of the target population. Quota-based samples are common practice in market research but are known to be second-best compared with the gold standard in the field, probability-based samples. A series of studies have compared the properties of the two sampling methods and found that quota-based sample techniques have improved markedly over the decades but still can fail to replicate the characteristics of the US population (e.g., MacInnis et al. 2018). In this case, it was the only cost-

[13] https://techcrunch.com/2019/12/04/prolific/

[14] We verified current prices for probability samples and found that a standard and well-respected online probability-based provider cost 10 times that of Prolific for an equivalent-sized panel.

effective option; care must be taken in interpreting the findings for the US population, especially any results that do not appear to be robust.

Across multiple iterations of the study, convenience samples were used in early iterations and the nationally representative, quota-balanced samples in later iterations. In total, the studies included 4,164 participants. Table C-9 shows each iteration of the study, and the source and number of participants in each.

Table C-9: Detail on Each Round of Research

#	Training Tool	Training Start	Training End	Test Tool	Test Start	Test End	Delay (Median)	Source	Total Starting Training	Training Done	Started Test	Completed Test	Completion Rate	Nat Rep?
1	Qualtrics	2/13/2021	2/13/2021	Qualtrics	2/13/2021	2/13/2021	None	Prolific	50	NA	NA	49	98%	No
2	Qualtrics	2/14/2021	2/14/2021	Qualtrics	2/14/2021	2/14/2021	None	Prolific	153	NA	NA	151	99%	No
3	Qualtrics	5/8/2021	5/8/2021	Qualtrics	5/8/2021	5/8/2021	None	Prolific	395	NA	NA	277	70%	No
4	Qualtrics	5/9/2021	5/13/2021	Qualtrics	5/13/2021	5/23/2021	1-13 days	Prolific	1,064	1032	879	725	68%	Yes
5P	Qualtrics	6/23/2021	6/28/2021	Rainloop	6/29/2021	7/17/2021	3-13 days	Prolific	438		249	190	43%	No
5D	Qualtrics	6/24/2021	6/27/2021	Rainloop	7/9/2021	7/12/2021	11-17 days	Dynata	564	457	352	86	15%	Yes (Attempted)
6A	Qualtrics	8/4/2021	8/6/2021	Rainloop	8/17/2021	8/20/2021	13 days	Prolific	750	729	561	533	71%	Yes
6B	Qualtrics	8/4/2021	8/6/2021	Rainloop	8/31/2021	9/8/2021	27 days	Prolific	750	692	464	434	58%	Yes

Center for Financial Security

School of Human Ecology
University of Wisconsin-Madison

1300 Linden Drive
Madison, WI 53706

608-890-0229
cfs@mailplus.wisc.edu
cfs.wisc.edu