Pwning Phishing Training Through Scientific Lure Crafting

•••

Dr. Christian Dameff, MD & Dr. Ariana Mirian, PhD

Black Hat 2025, Human Factors Track

Who are we?

- Associate professor @ UCSD
- Co-director @ UCSD Center for Healthcare Cybersecurity

- Security researcher focused on Internet measurement/security
- Currently @ Censys, Previously PhD @ UCSD

Agenda

- Background & Motivation
- Study Setup, Design, & Methods
- Lessons Learned (and what that means for users)
- Summary

Audience poll: Does user phishing training work?

Background + Motivation

...

Phishing Training works...right?

- Many organizations (including ours) perform trainings
  - Annual cybersecurity awareness trainings
  - Simulated phishing tests (embedded trainings)

- Teach a person to spot a phish, and they are trained for life
  - “Human firewalls”

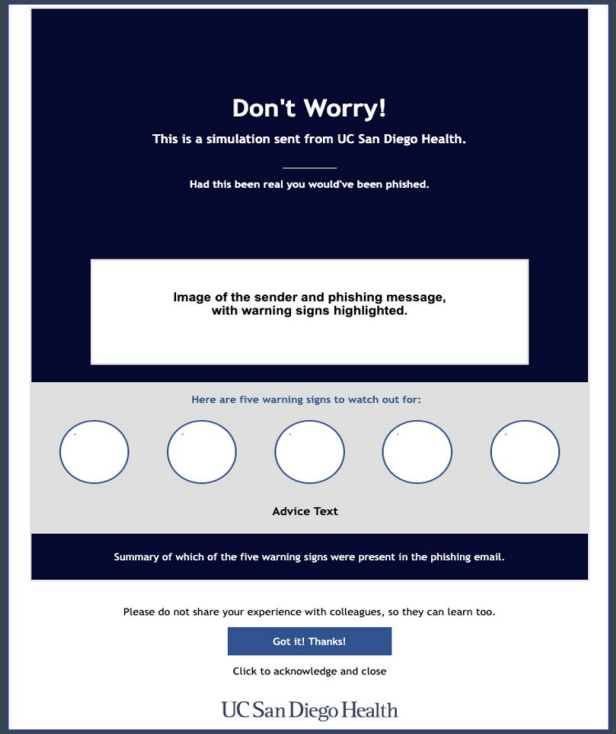
Background

- Much prior research is in favor of anti-phishing training
  - i.e : [Jampen et al. 2020]
  - Often lab studies
- Some recent studies that show opposite results
  - I.e : [Lain et al. 2022]
  - Increasingly real world studies with actual users

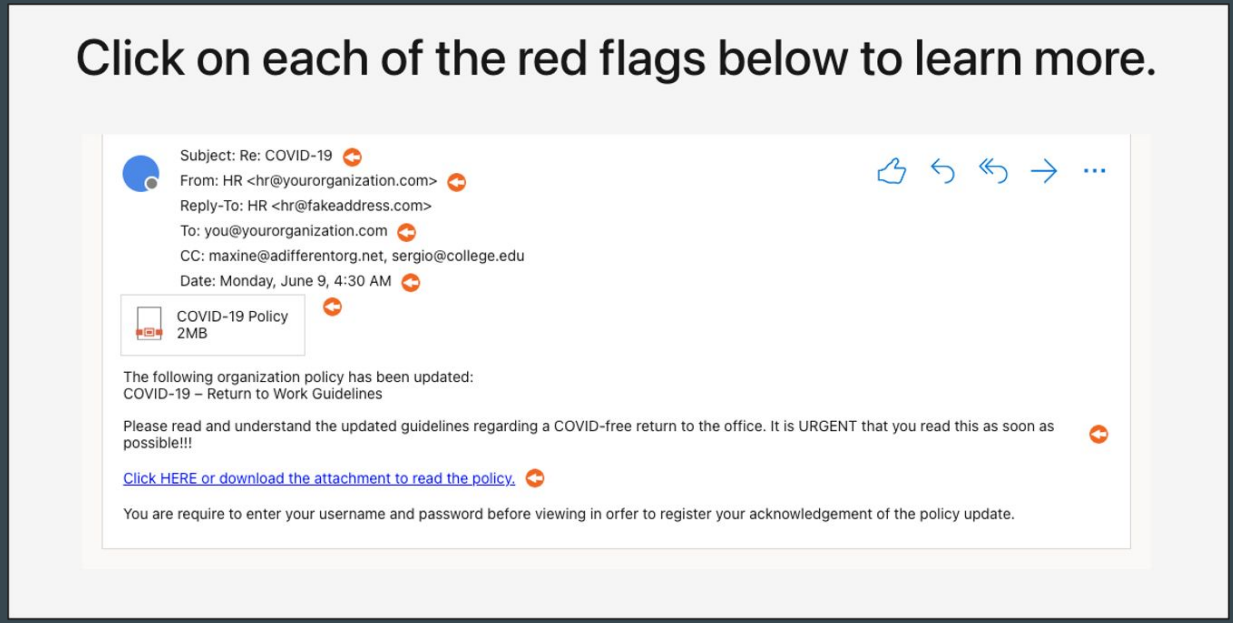
- **Problem: How do we reconcile these conflicting studies?**

Underlying research question:
What is the best modality for anti-phishing training?

Many different modalities – which to focus on?



Static



Interactive

Let's Treat Security Research like Medical Research

Medical Outcomes

Security Outcomes

Let's Treat Security Research like Medical Research

The Drug Development Process

Step 1
Discovery and Development

Discovery and Development
Research for a new drug begins in the laboratory.
More Information

Step 2
Preclinical Research

Preclinical Research
Drugs undergo laboratory and animal testing to answer basic questions about safety.
More Information

Step 3
Clinical Research

Clinical Research
Drugs are tested on people to make sure they are safe and effective.
More Information

Step 4
FDA Review

FDA Review
FDA review teams thoroughly examine all of the submitted data related to the drug or device and make a decision to approve or not to approve it.
More Information

Step 5
FDA Post-Market Safety Monitoring

FDA Post-Market Safety Monitoring
FDA monitors all drug and device safety once products are available for use by the public.
More Information

Medical Outcomes

Security Outcomes

Let's Treat Security Research like Medical Research

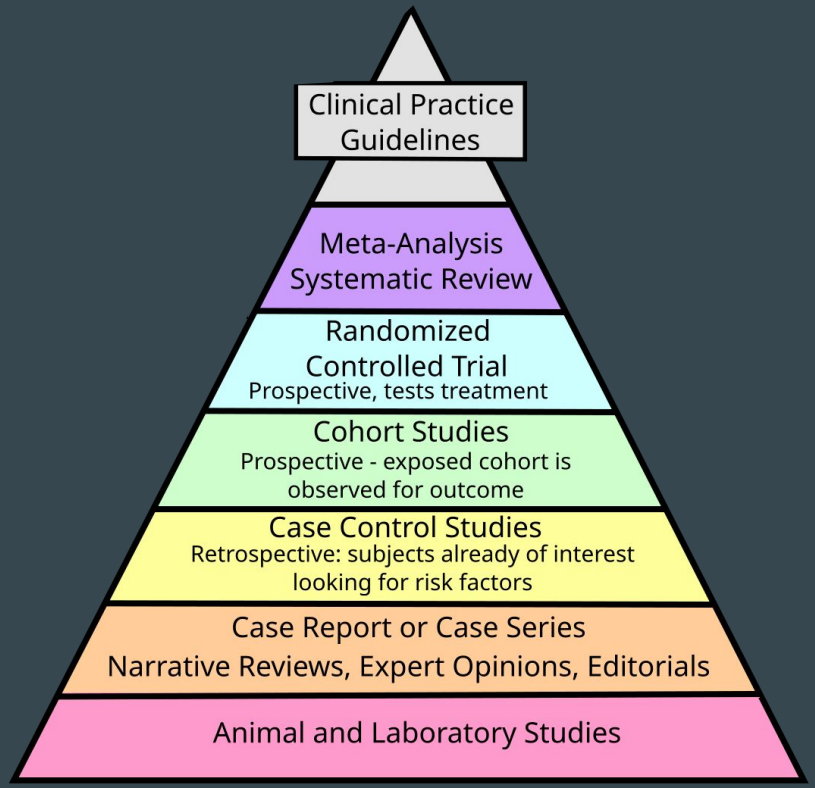
- Evidence based cybersecurity should be the norm.
  - Bloodletting & mercury = bad

- Instead of spending millions of dollars AND hours on ineffective solutions, let's find the EFFECTIVE ones with science.

Methodology

...

Not all evidence is equal



Randomized 19,000+ Employees into 5 Groups

- Control (no training)
- Generic static
- Generic interactive
- Contextual static
- Contextual interactive

The 8 month experiment

- Deployed monthly simulated phishing tests
  - If user clicked, they got one of four trainings
  - Control group failure led to 404 page
- Users got 1/10 lures
- Collected:
  - User failure rates
  - Training engagement (ie. time on page)
  - Time since last annual cybersecurity training
  - And additional data

Phishing Lure
Outlook Pwd
Login Account
Open Enroll
Shared Doc (Microsoft)
OneDrive Medical
Docusign
Building Evac
Traffic Ticket
Dress Code
Vacation Policy

Lure example

Hello,

The IT department has found that your logon password has been stolen by a hacker! We need you to update your password with our database or it will be disabled, preventing you from accessing the system. Please go to the URL below and enter your current username and password before your access is revoked:

Click here to reset your password

Thank you in advance for your cooperation.
IT Support

Lessons Learned
(and what it means for users)
…

Lesson #1: We Can Pwn Users with Scientific Lure Crafting

...

Lesson #1: we can pwn users with scientific lure crafting

Phishing Lure	# of Users	Avg Failure Rate
Outlook Pwd	4,931	1.82%
Login Account	12,720	1.85%
Open Enroll	14,691	7.62%
Shared Doc (Microsoft)	15,683	8.99%
OneDrive Medical	18,438	9.20%
Docusign	23,526	9.63%
Building Evac	17,359	10.33%
Traffic Ticket	17,676	18.60%
Dress Code	4,954	27.65%
Vacation Policy	17,923	30.80%

Top Tier Lure Example

Dear %FIRSTNAME%,

Please be advised that as part of our ongoing review process, we plan to institute several fundamental changes to our dress code. Please view these changes by visiting the Human Resources website.

This policy will go into effect 30 days from the receipt of this notice. It is up to you to know and comply with this change in dress code.

Any staff member who does not meet the attire or grooming standards set by his or her department will be subject to disciplinery action and may be asked to leave the premises to change clothing. Hourly paid staff members will not be compensated for any work time missed because of failure to comply with designated workplace attire and grooming standards.

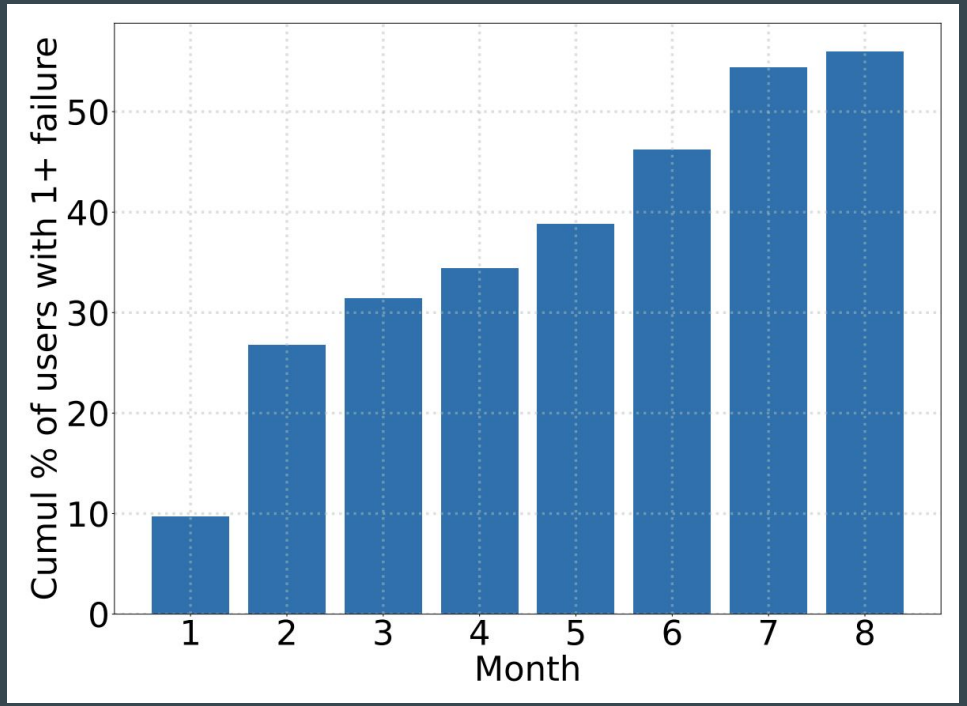
Regards,
Human Resources
UC San Diego Health

Lesson #1: we can pwn users with scientific lure crafting

- Whoever controls the lures, controls the failure rate!

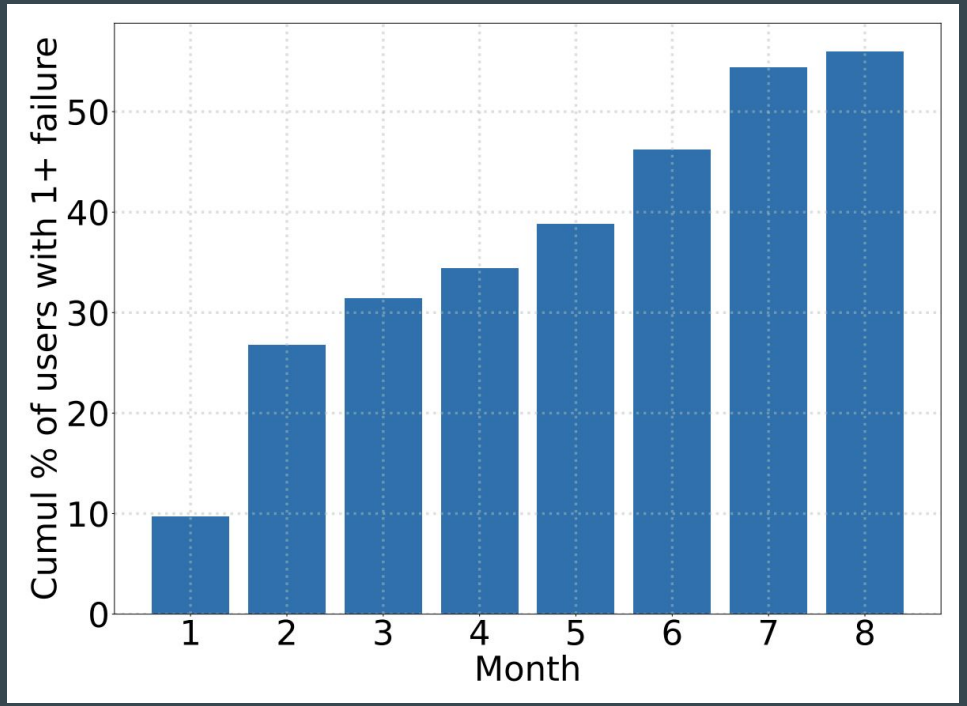
Lesson #1: we can pwn users with scientific lure crafting

- Whoever controls the lures, controls the failure rate!
- On a long enough time frame, most people are pwned.



Lesson #1: we can pwn users with scientific lure crafting

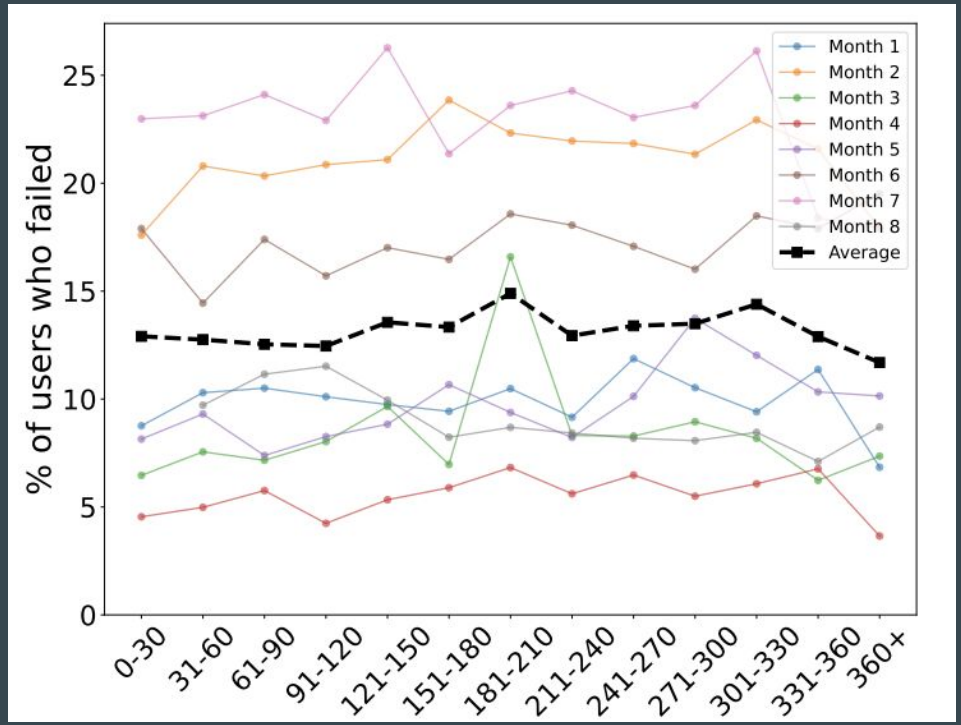
- Whoever controls the lures, controls the failure rate!
- On a long enough time frame, most people are pwned.
- We need to stop punishing employees for failing phish.



Lesson #2: Training not efficacious (for these modalities/deployment)

...

Lesson #2: training not largely efficacious (in these modalities/deployment)



Annual cybersecurity training has no observable benefit

Lesson #2: training not largely efficacious (in these modalities/deployment)

Overall average improvement over control for monthly embedded training was….1.7%

Lesson #2: training not largely efficacious (in these modalities/deployment)

Overall average improvement over control for monthly embedded training was....1.7%

Lesson #2: training not largely efficacious (in these modalities/deployment)

Overall average improvement over control for monthly embedded training was....1.7%

Lesson #3: People Don’t Spend Time on Anti-Phishing Trainings

…

Lesson #3: people don't spend time on training

	Generic		Contextual	
Statistic	Static	Interactive	Static	Interactive
Sessions w/ 0 sec	39.7%	51.3%	37.3%	44.3%
25th percentile	0 sec	0 sec	0 sec	0 sec
50th percentile	7 sec	0 sec	10 sec	6 sec
75th percentile	19 sec	24 sec	27 sec	48 sec
90th percentile	34 sec	70 sec	52 sec	101 sec

- We measured how much time folks are spending on training

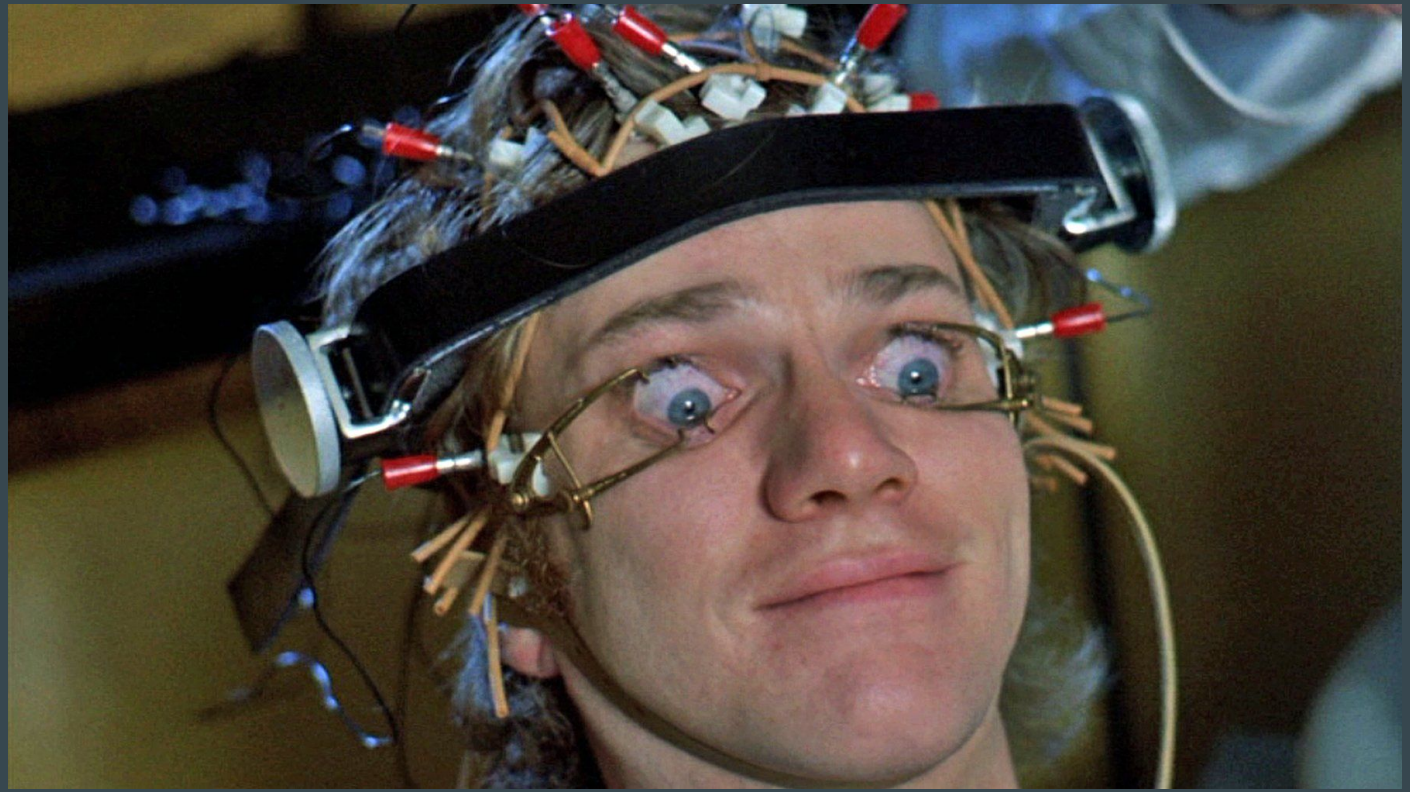
Lesson #3: people don't spend time on training

- For the people who did spend time on training, there were different outcomes

- Static trainers did **worse** , interactive trainers did **better**

- Overall numbers were really low, so hard to generalize

Lesson #3: people don’t spend time on training



Is all of this focus on training worth the outcomes?

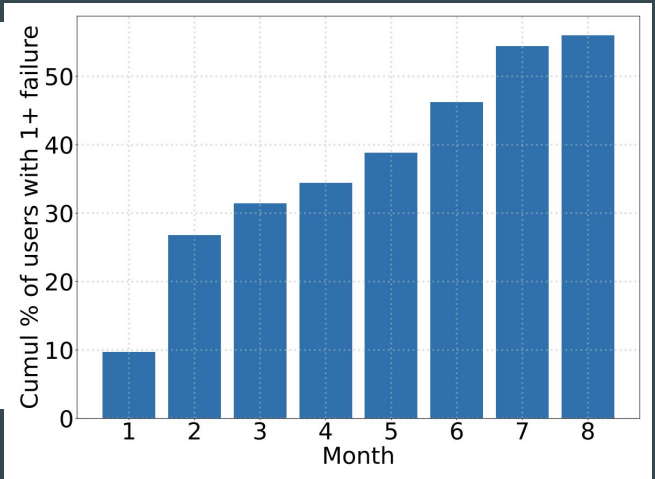
...

We know:

Phishing Lure	# of Users	Avg Failure Rate
Outlook Pwd	4,931	1.82%
Login Account	12,720	1.85%
Open Enroll	14,691	7.62%
Shared Doc (Microsoft)	15,683	8.99%
OneDrive Medical	18,438	9.20%
Docusign	23,526	9.63%
Building Evac	17,359	10.33%
Traffic Ticket	17,676	18.60%
Dress Code	4,954	27.65%
Vacation Policy	17,923	30.80%

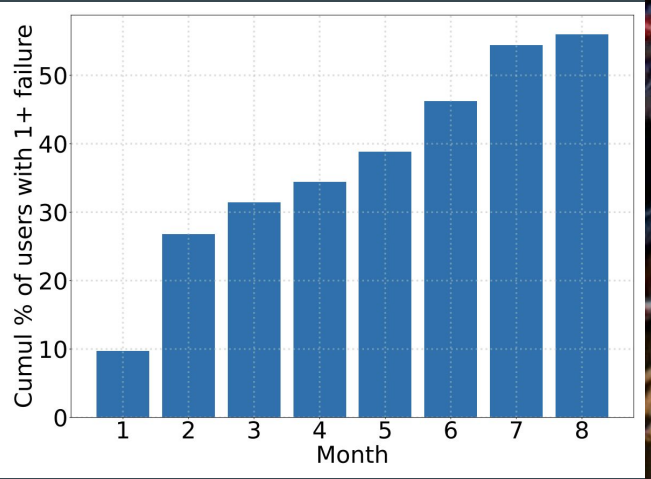
We know:

Phishing Lure	# of Users	Avg Failure Rate
Outlook Pwd	4,931	1.82%
Login Account	12,720	1.85%
Open Enroll	14,691	7.62%
Shared Doc (Microsoft)	15,683	8.99%
OneDrive Medical	18,438	9.20%
Docusign	23,526	9.63%
Building Evac	17,359	10.33%
Traffic Ticket	17,676	18.60%
Dress Code	4,954	27.65%
Vacation Policy	17,923	30.80%



We know:

Phishing Lure	# of Users	Avg Failure Rate
Outlook Pwd	4,931	1.82%
Login Account	12,720	1.85%
Open Enroll	14,691	7.62%
Shared Doc (Microsoft)	15,683	8.99%
OneDrive Medical	18,438	9.20%
Docusign	23,526	9.63%
Building Evac	17,359	10.33%
Traffic Ticket	17,676	18.60%
Dress Code	4,954	27.65%
Vacation Policy	17,923	30.80%



is all of this focus on training worth the outcome?

- We CAN find the “right” training
- How much time/effort/money will it take us?
- How much would be erased with a slightly different lure?

is all of this focus on training worth the outcome?

- We CAN find the “right” training
- How much time/effort/money will it take us?
- How much would be erased with a slightly different lure?

What if we put energy and resources elsewhere?

We need to empirically measure these outcomes, and share the data, to better security.

...

Let's Treat Security Research like Medical Research

The Drug Development Process

Step 1
Discovery and Development

Discovery and Development
Research for a new drug begins in the laboratory.
More Information

Step 2
Preclinical Research

Preclinical Research
Drugs undergo laboratory and animal testing to answer basic questions about safety.
More Information

Step 3
Clinical Research

Clinical Research
Drugs are tested on people to make sure they are safe and effective.
More Information

Step 4
FDA Review

FDA Review
FDA review teams thoroughly examine all of the submitted data related to the drug or device and make a decision to approve or not to approve it.
More Information

Step 5
FDA Post-Market Safety Monitoring

FDA Post-Market Safety Monitoring
FDA monitors all drug and device safety once products are available for use by the public.
More Information

Medical Outcomes

Security Outcomes

broaden data sharing

- Back-up claims with data
- Should vendors be the collector, disseminator, and analyzer of data?
- We don’t need to be an expert, but let’s get data in the hands of the RIGHT people

Summary

...

In summary

- Lesson #1: we can pwn users with scientific lure crafting
- Lesson #2: trainings (as deployed) are not efficacious
- Lesson #3: people don’t spend time on training

In summary

- Recommendation #1: Let's find the more efficacious places to put time and energy
- Recommendation #2: Empirically analyze security outcomes. Always.

Audience poll: Does user phishing training work?

Understanding the Efficacy of Phishing Training in Practice

Grant Ho[◇†] Ariana Mirian[◁†] Elisa Luo[†] Khang Tong[⋆‡] Euyhyun Lee[⋆‡]
Lin Liu[⋆‡] Christopher A. Longhurst[⋆] Christian Dameff[⋆] Stefan Savage[†] Geoffrey M. Voelker[†]

[†]UC San Diego [◇]University of Chicago [⋆]UC San Diego Health

***Abstract*—This paper empirically evaluates the efficacy of two ubiquitous forms of enterprise security training: annual cybersecurity awareness training and embedded anti-phishing training exercises. Specifically, our work analyzes the results of an 8-month randomized controlled experiment involving ten simulated phishing campaigns sent to over 19,500 employees at a large healthcare organization. Our results suggest that**

covering over 133M health records, and 460 associated ransomware incidents (more than one per day) [2], [11].

Absent an effective technical defense, organizations have turned to security training as a means to staunch the bleeding. Our own institution admonishes each of us to "Be a Human Firewall" — to identify and resist enticements to click on suspicious email-borne links. Indeed, in many sec-

https://arianamirian.com/docs/ieee-25.pdf

Thank you!

@quaddi@gmail.com
@cyberhealth.ucsd.edu
@cdameff.bsky.social
@cdameff

@arianamirian28@gmail.com
@arianamirian.com
@arianamirian.bsky.social
@arianamirian