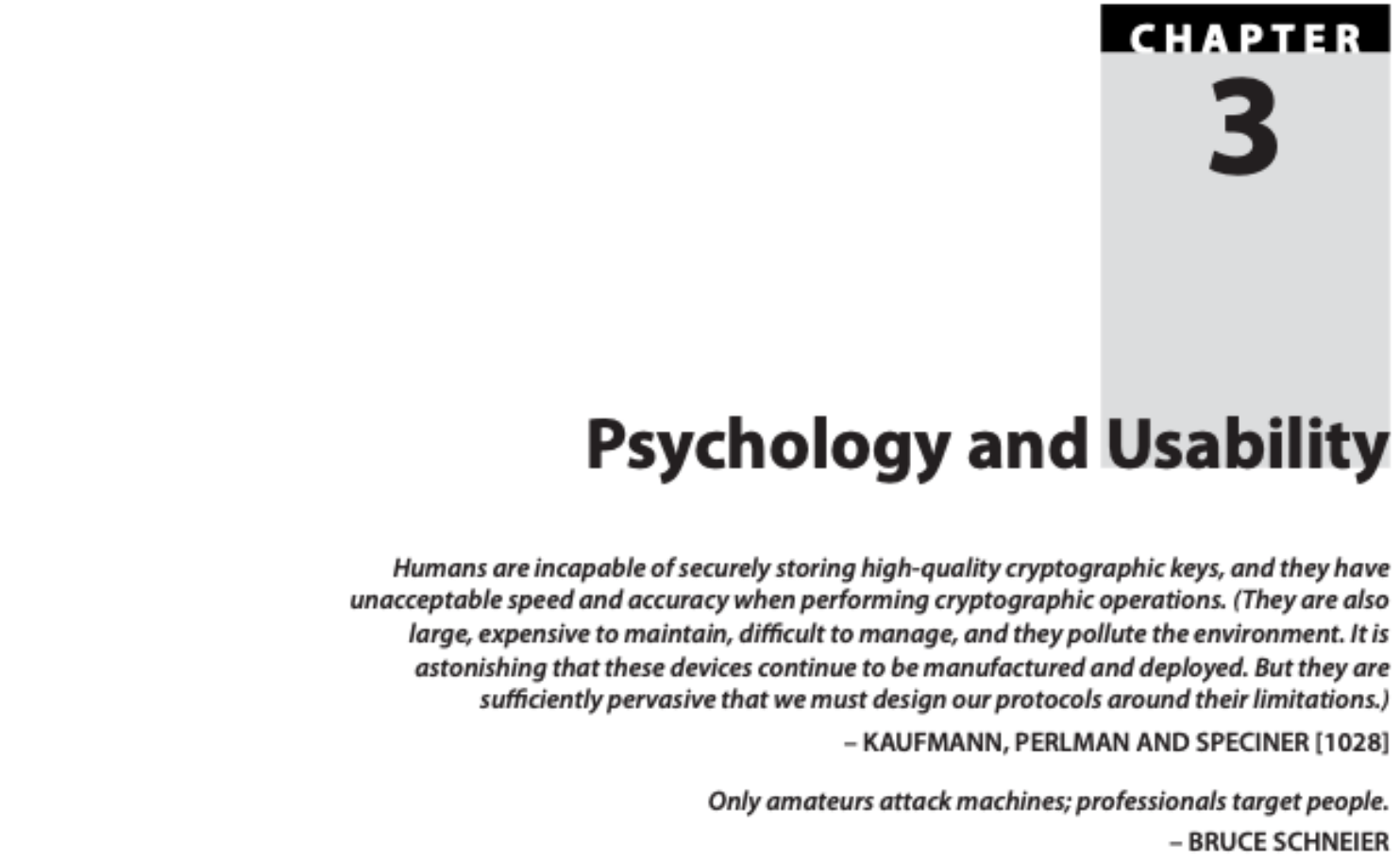
Security Engineering

INFR11208 (UG4) // NFR11228 (MSc)

Daniel W. Woods* **and Jingjie Li**
Email: daniel.woods@ed.ac.uk **and** jingjie.li@ed.ac.uk

Chapter 8, Security Engineering



CHAPTER

3

Psychology and Usability

Humans are incapable of securely storing high-quality cryptographic keys, and they have unacceptable speed and accuracy when performing cryptographic operations. (They are also large, expensive to maintain, difficult to manage, and they pollute the environment. It is astonishing that these devices continue to be manufactured and deployed. But they are sufficiently pervasive that we must design our protocols around their limitations.)

– KAUFMANN, PERLMAN AND SPECINER [1028]

Only amateurs attack machines; professionals target people.

– BRUCE SCHNEIER

https://www.cl.cam.ac.uk/archive/rja14/Papers/SEv3-ch3.pdf

Why psychology

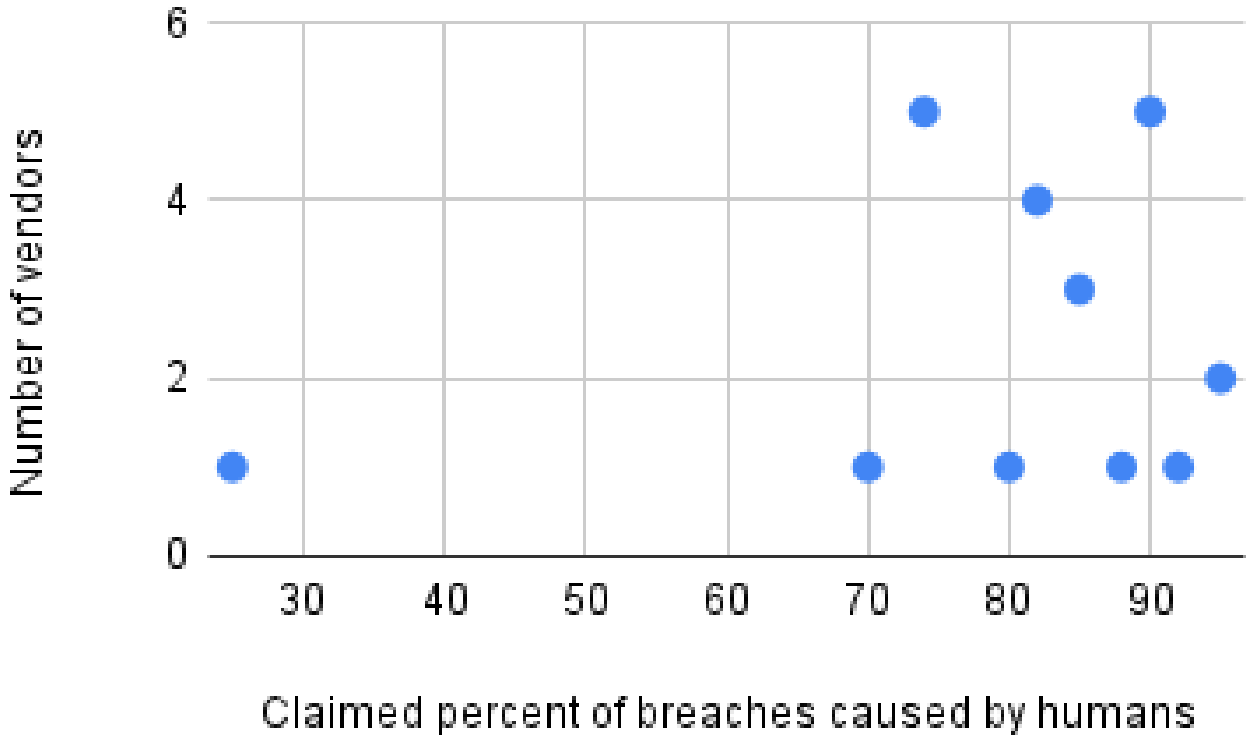
"**LO3:** Compare and synthesise the perspectives of different system stakeholders and threat actors, using economic and psychological viewpoints as well as technical ones."

Summary

1. **What kind of user errors lead to losses**
   - Social engineering
   - Phishing
2. **What psychological dynamics do they exploit**
   - Prospect theory

3. **How can we design systems to account for it**
   - Training
   - Warnings

How many attacks involve human error?



Source: Hielscher, J., Schöps, M., Opdenbusch, J., Reichmann, F., Gutfleisch, M., Marky, K., & Parkin, S. (2024, December). Selling Satisfaction: A Qualitative Analysis of Cybersecurity Awareness Vendors' Promises. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security* (pp. 2666-2680).

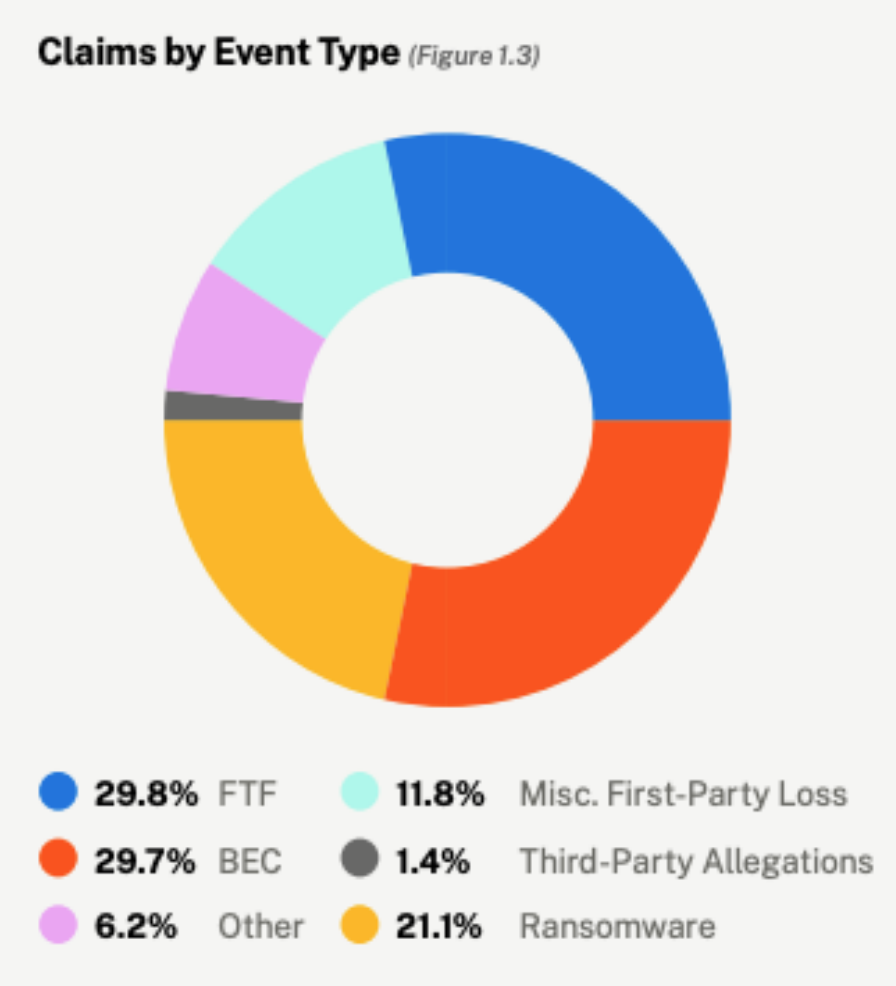
Recall: Most crimes are authorized by the victim

2024 CRIME TYPES *continued*

BY COMPLAINT LOSS

Crime Type	Loss	
Investment	$6,570,639,864	→ Social engineer individuals
Business Email Compromise	$2,770,151,146	→ Social engineer business, sometimes via hack
Tech Support	$1,464,755,976	→ Social engineer individuals
Personal Data Breach	$1,453,296,303	→ Mostly remote hacking
Non-Payment/Non-Delivery	$785,436,888	→ Defraud ecommerce buyers
Confidence/Romance	$672,009,052	} Social engineer individuals
Government Impersonation	$405,624,084	
Data Breach	$364,855,818	→ Mostly remote hacking
Other	$280,278,325	
Employment	$264,223,271	→ Weird but not remote hacking
Credit Card/Check Fraud	$199,889,841	→ Auth failures, but not via remote hacking
Identity Theft	$174,354,745	→ Auth failures at banks, using data from hacking
Real Estate	$173,586,820	→ Social engineer renters/house buyers

Corporate incidents



Types of "Human Errors"

- FTF due to two types:
  - Spoofed emails
  - Failure to adhere to processes
- BEC can result from:
  - Phishing
  - Infostealers
  - Setting weak passwords
- Ransomware (+ data breach) can result from:
  - Phishing + Infostealers
  - Social engineering
  - Organizational dynamics

Source: Coalition. 2025 Cyber Claims Report.

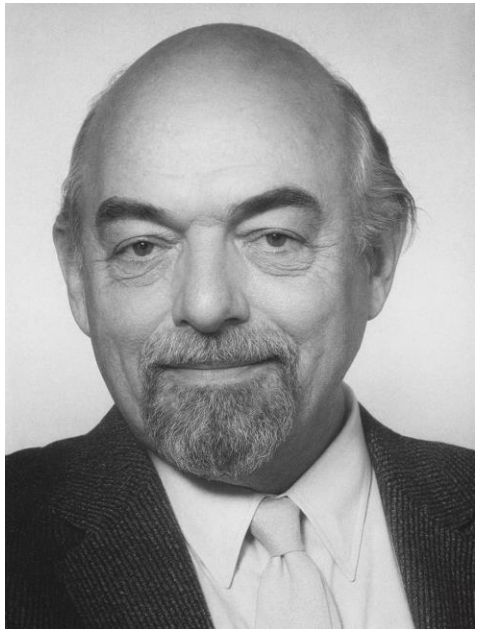
The SRK Model

A taxonomy of errors that impact cybersecurity

Psychology of safety and security

Rough Taxonomy of "Errors"

- Errors aren't random or due to "stupidity"
- Errors arise at different levels of the ‘stack’ (Rasmussen, 1983)
  - Deal with novel problems in a conscious way
  - Frequently encountered problems are dealt with using rules we evolve, and are partly automatic
  - Over time, the rules give way to skill
- Under pressure, humans regress from knowledge → rule → skill-based control
- Automatising routine actions leads to:
  - absent-minded slips
  - following a wrong rule
- There are also systematic limits to rationality – ‘heuristics and biases’, as well as social psychology



Jens Rasmussen

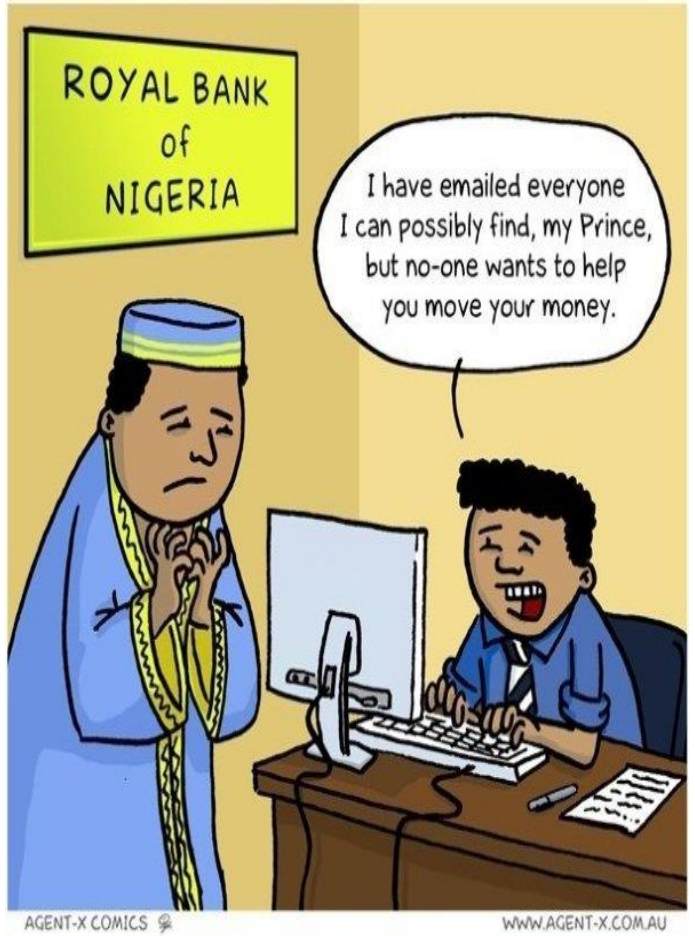
Rasmussen's Skill, Rule and Knowledge (SRK) Model

	Skills-based behaviour	Rules-based behaviour	Knowledge-based behaviour
Representation	Automatic, sensorimotor control	Stored rules and procedures	Mental models of the system
Type of thinking	Unconscious, fast	Depends	Conscious, slow
Format	Continuous, analog	Discrete, symbolic	Abstract, causal
Examples	Checking URLs and email addresses	Patch management via severity	Engineering a system

Attacking knowledge-based behaviors

- Advance payment fraud
  - Fraudsters present mark with an opportunity for:
    - Fantastic investment returns
    - Rental property
    - A Nigerian prince's frozen funds
- CEO Fraud
  - Impersonate CEO and tell mark to send funds asap
    - Time pressure to prevent *too much* thinking

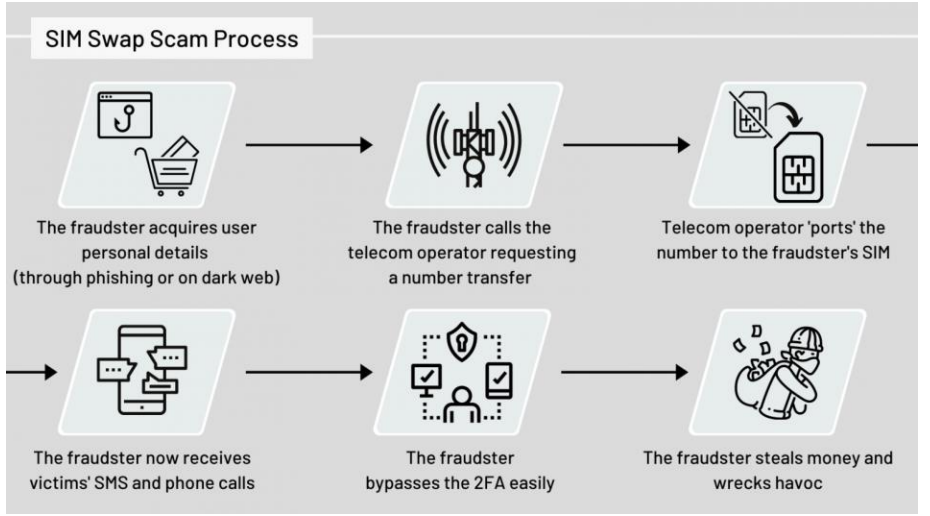
Knowledge-based attacks exploit human reasoning under **uncertainty** and **novelty.** Hard to pull off.



Attacking rule-based behaviors

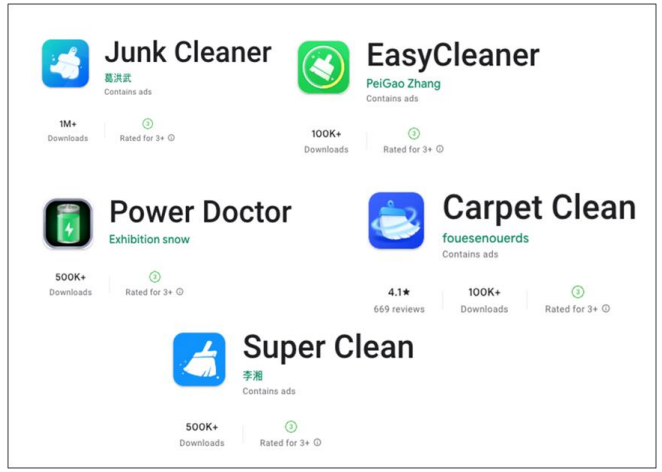
- Customer support procedures
  - Telling phone company "you" lost your phone and need a new SIM card
  - Telling IT help desk you've lost access to account
- Tell finance department that their vendor's bank details have changed
  - 2nd most costly for of cybercrime according to FBI
  - Often combined with spoofing the email to exploit skill-based behaviour (spotting URL)
- Malware in app store

Exploit rule-based behaviours by forcing victim to **apply rules inappropriately** or **overload procedures**, leading to mistakes.



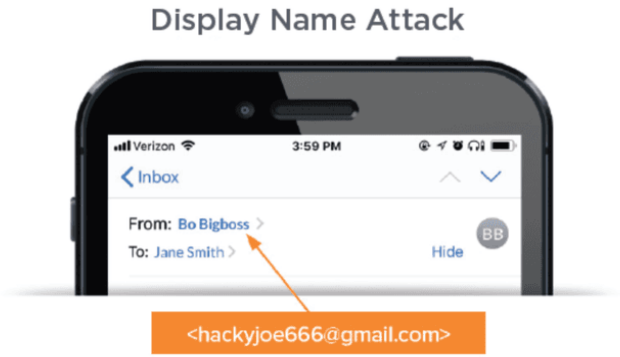
MGM Resorts hack: How attackers hit the jackpot with service desk social engineering

Marcus White | Last updated on **November 12, 2025**



Attacking skill-based behaviors

- Exploit motor skill slips
  - Typo squatting: gogle.com
- Exploit visual recognition limitations
  - Foreign characters: goo**o**gle.com --> exploits visual recognition
  - Subdomain manipulation: office365.store.com
  - Displayname: <Your boss, boss@criminal.com>
- Exploit habituation
  - Pop up with X that hyper links to malware
  - MFA alert fatigue



Exploit skill-based behaviours by **mimicking familiar sensory cues** to trigger **automatic reflexes**, leading to absent-minded slips

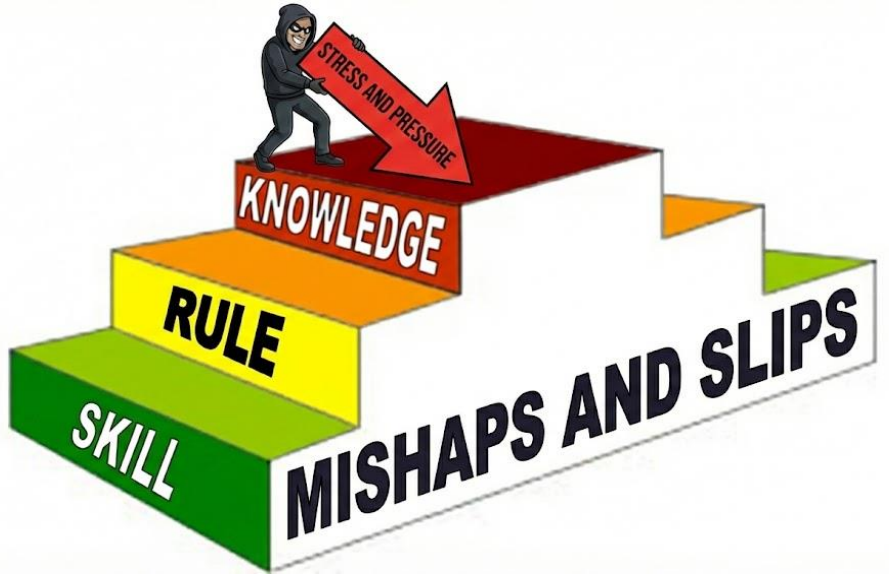
What's the difference between safety and security?

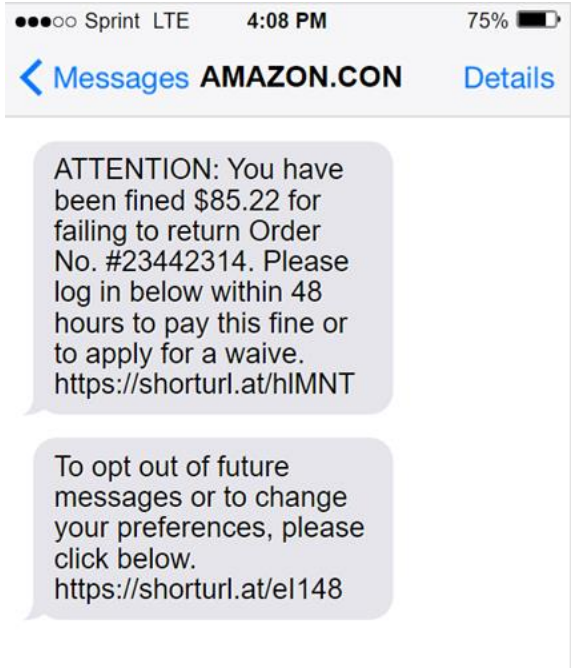
Adversaries create situations that intensify errors

Attacks exploit urgency

- **Key Idea:** Under pressure, humans regress from knowledge --> rule --> skill-based control"
  - In safety, stress is an unfortunate byproduct
  - In security, stress is manufactured.
- How threat actors manufacture stress
  - Time pressure
  - Authority

Goal: Disable "Knowledge-based" reasoning (System 2) and force a "Rule-based" or "Skill-based" reaction (System 1)





Many attacks exploit authority figures

- **Authority:** Stanley Milgram showed that over 60% of all subjects would inflict a potentially fatal shock on a ‘student’ if ordered to do so by a ‘teacher’
- **Hierarchies:** In 1966, researchers found that 95% of nurses were willing to administer a lethal drug dose just because a voice on the phone claimed to be a doctor
- Deference to specialists
  - For IT and fraud, people assume "it's too complex" to understand and trust the specialist

2024 CRIME TYPES *continued*

BY COMPLAINT LOSS

Crime Type	Loss
Investment	$6,570,639,864
Business Email Compromise	$2,770,151,146
Tech Support	$1,464,755,976
Personal Data Breach	$1,453,296,303
Non-Payment/Non-Delivery	$785,436,888
Confidence/Romance	$672,009,052
Government Impersonation	$405,624,084
Data Breach	$364,855,818
Other	$280,278,325

Many attacks exploit love

US mother gets call from 'kidnapped daughter' – but it's really an AI scam

Jennifer DeStefano tells US Senate about dangers of artificial technology after receiving phone call from scammers sounding exactly like her daughter

Erum Salam

Wed 14 Jun 2023 20.05 BST

2024 CRIME TYPES *continued*

BY COMPLAINT LOSS

Crime Type	Loss
Investment	$6,570,639,864
Business Email Compromise	$2,770,151,146
Tech Support	$1,464,755,976
Personal Data Breach	$1,453,296,303
Non-Payment/Non-Delivery	$785,436,888
Confidence/Romance	$672,009,052
Government Impersonation	$405,624,084
Data Breach	$364,855,818
Other	$280,278,325
Employment	$264,223,271
Credit Card/Check Fraud	$199,889,841
Identity Theft	$174,354,745
Real Estate	$173,586,820

Many attacks exploit curiosity

Conferences > 2016 IEEE Symposium on Securi...

Users Really Do Plug in USB Drives They Find

Publisher: IEEE

Cite This | PDF

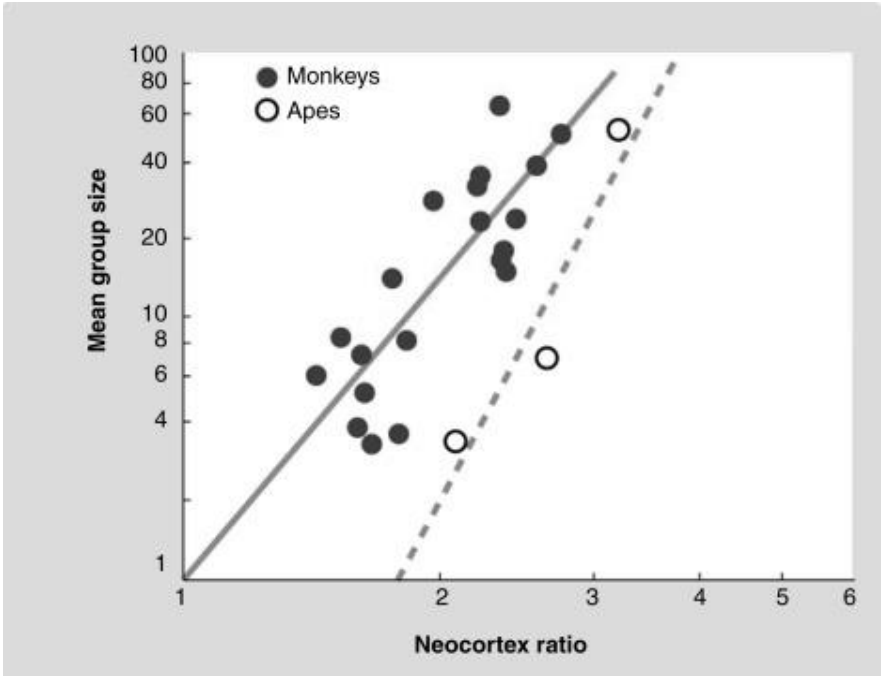
Matthew Tischer ; Zakir Durumeric ; Sam Foster ; Sunny Duan ; Alec Mori ; Elie Bursztein

All Authors

- Dropped 300 USBs across campus
- Users picked up, plugged in, and clicked on files in 48% of the drives we dropped.
- They did so quickly: the first drive was connected in under six minutes

The social brain hypothesis

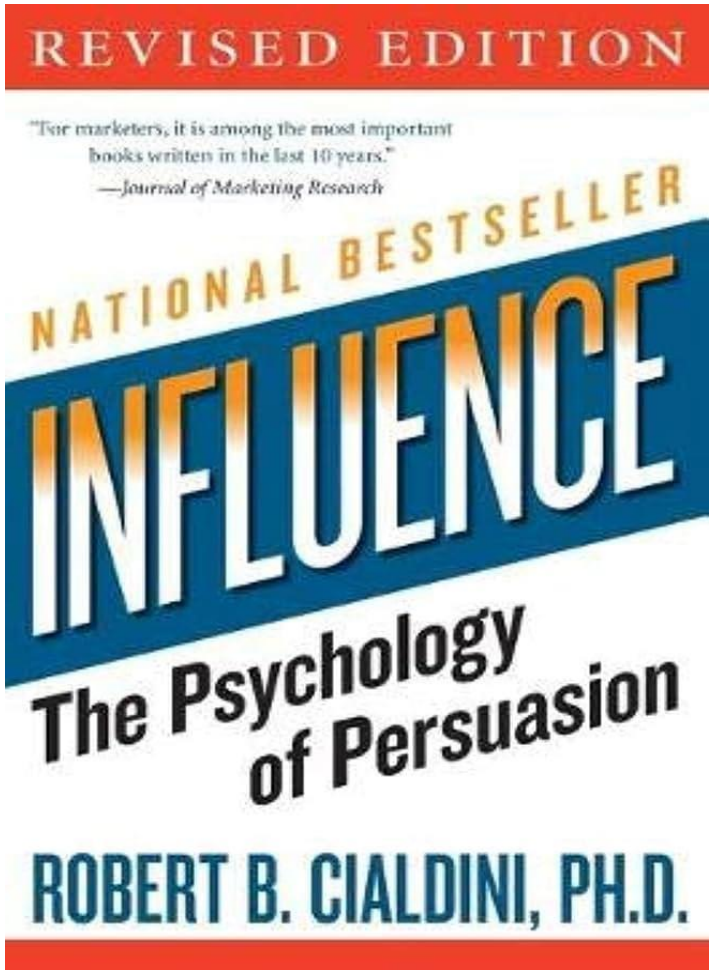
- Old view: we got smart to make better tools
  - Archaeology: we got smart first!
- New view: when Africa dried out 1.5m years ago, we started living in bigger groups
  - Primate brain size correlates well with group size
  - Social aspect: big brains track more relationships
  - Machiavellian aspect: if you're better at deception, and at detecting deception in others, you're more likely to have descendants



The "science" of deception

Cialdini’s 6 Principles of Persuasion

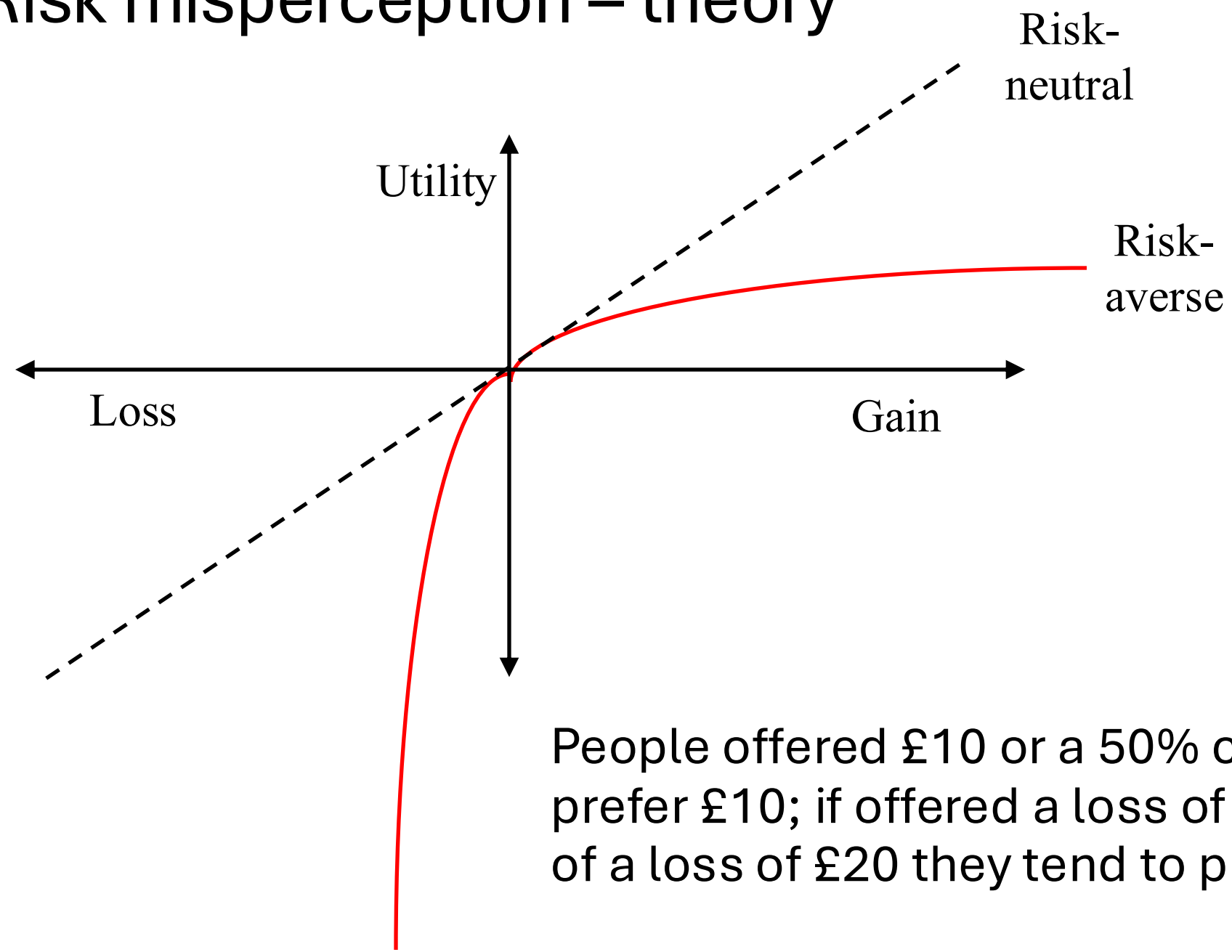
- **Reciprocity:** "Here is a free industry whitepaper."
  - Victim feels obliged to "register" (hand over credentials) to download it.
- **Scarcity:** "24 hours until account deletion."
  - Victim clicks the malicious link to "save" their data.
- **Authority:** "This is the CEO. Wire these funds immediately."
  - Victim bypasses verification rules to obey a superior.
- **Consistency:** "You signed our new remote work policy."
  - "So you must install this 'compliance tool' (malware) to match your agreement."
- **Liking:** "I love your LinkedIn posts! We went to the same college."
  - Attacker builds rapport to send a weaponized resume later.
- **Social Proof:** "90% of your colleagues have already completed this survey."
  - Victim clicks to avoid being the "odd one out."



Behavioural economics

Even without an adversary, humans show bounded rationality

Risk misperception – theory



People offered £10 or a 50% chance of £20 usually prefer £10; if offered a loss of £10 or a 50% chance of a loss of £20 they tend to prefer the latter!

Decisions are heavily influenced by framing

- The ‘Asian disease problem’ where the subject is making decisions on vaccination. Two options put to subjects. First:

  A: “200,000 lives will be saved”

  B: “with p=1/3, 600,000 saved; but p=2/3 none saved”

  - Here 72% choose A over B!
- Second option is

  C: “400,000 will die”

  D: “with p =1/3, no-one will die, p=2/3, 600,000 die”

  - Here 78% prefer D over C!
- This is also why marketers talk ‘discount’ or ‘saving’ – and fraudsters know that **people facing losses take more risks**

Security implications

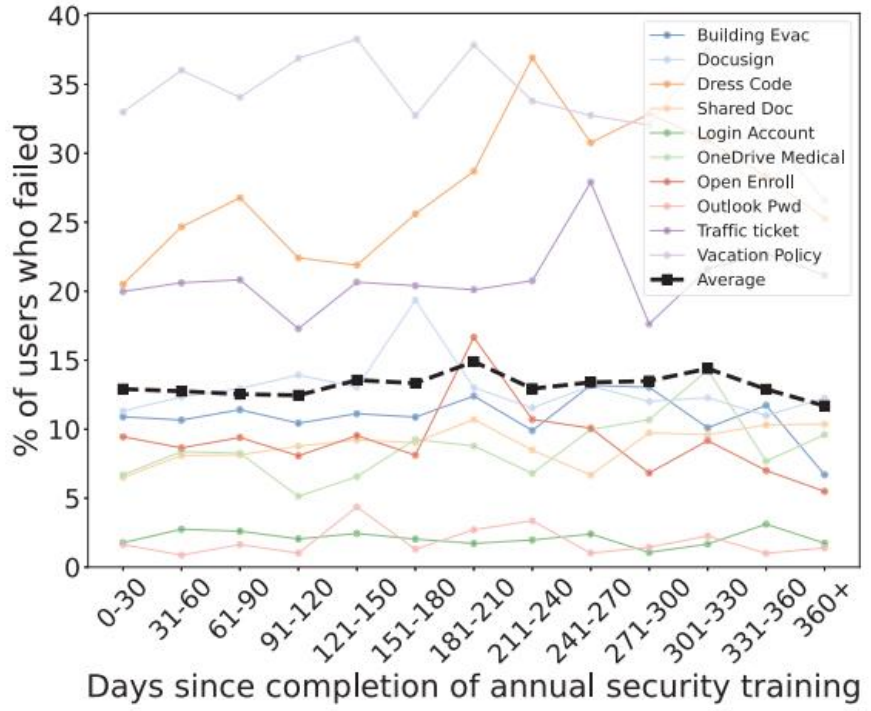
- Ransomware victims
  - Rebuilding from backups or letting them leak data feels like a guaranteed loss
  - Paying the ransom has a chance of zero loss
- Data breach victims and reporting
  - Once a breach starts, IT teams often try to "fix it quietly" rather than report it immediately
- Fraud victims
  - Frame initial investment as guaranteed return
    - Risk averse prefer guaranteed return from fraudster to uncertainty of public markets
  - After £50k of payments, they face "Asian disease"
    - Accept that it's a scam and lose £50k
    - Pay an extra £10k to "unfreeze" funds and have a non-zero chance of losing nothing

How to organize defenses with security in mind

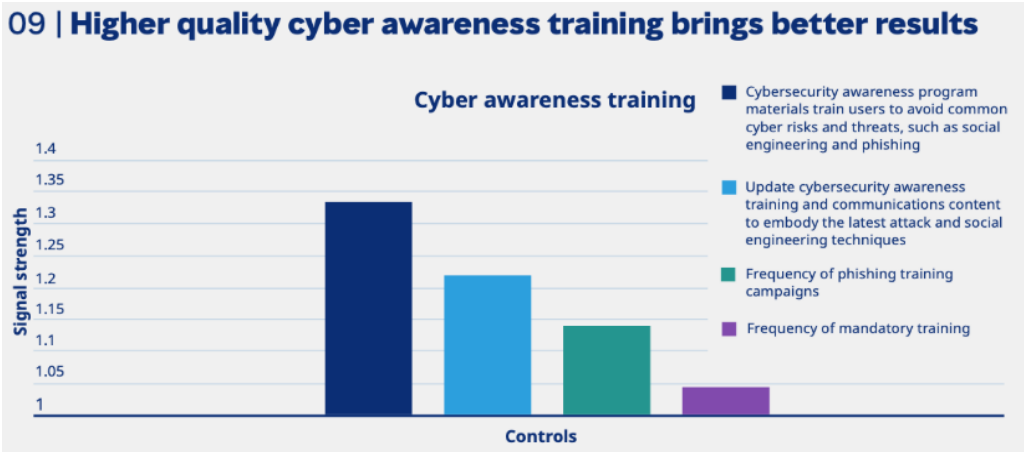
Psychology-aware security engineering

Training

- Most organizations train users annually
  - Often a compliance requirement
- Marginal impact on phishing failure
  - When you took the training has no impact
  - Random trial shows embedded training reduces failure rate by only 2%
  - 10% of users failed in the first month, but that rose to over half of users after 8 months
- However, insurance data shows security awareness training works (see Marsh)
  - Frequency of training doesn't matter much
  - Training to avoid common cyber risks and threats is correlated with lower incidents!
    - Recall SKR model!
- Training as similar to vaccines
  - See my article

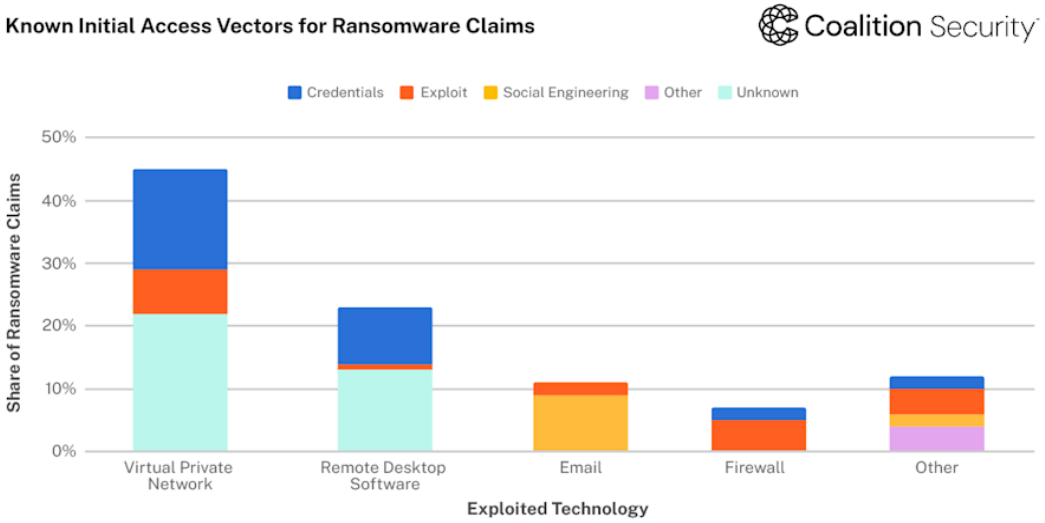


Source: Ho, Grant et al. "Understanding the efficacy of phishing training in practice." In *2025 IEEE Symposium on Security and Privacy (SP)*, pp. 37-54. IEEE, 2025.



Users' mental models

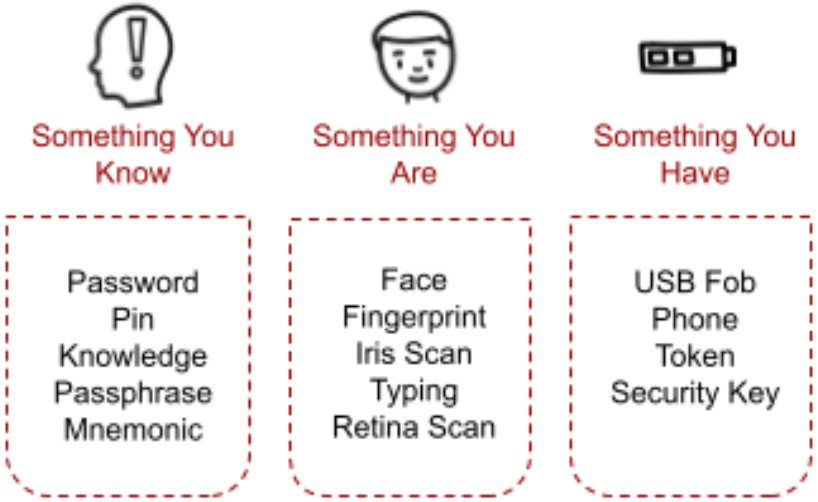
- Explore how your users see the problem – the 'folk beliefs'
  - 'perimeterisation' as a bad mental model of enterprise networks
  - Who are 'hackers' and who do they target (Wash, 2010)
  - Passwords as closely guarded one-time secrets, instead of strings that are constantly re-used, shared and held by insecure 3rd parties
  - Fraud as dumb and obvious (e.g. Nigerian prince scams)
- People are more likely to follow security advice consistent with their mental model



	Graffiti	*Burglar*	*Big Fish*	*Contractor*
# Subjects	8	13	9	3
Identity of hacker(s)	Young technical geek	Some criminal	Professional criminal hackers	Young technical geek
Level of organization	Solo, or to impress friends	Unspecified	Part of a criminal organization	Solo, but a contractor for criminals
Reason for break-ins	Cause mischief	Look for financial and personal information	Look for financial and personal information	Look for financial and personal information
Effects of break-ins	Lots of computer problems; requires reinstall	Possible harm to computer; exposure of personal information	No harm to computer; exposure of personal information	Exposure of personal information
Target(s)	Anyone; doesn't matter	Opportunistic; could be me	Not me; only looking for rich or important people	Not me; looking for large databases of info
Am I a target?	Possibly	Possibly	No	No

Table 2: Summary of folk models about hackers, organized by model features

User authentication



- Something you know
  - Passwords are cheap but…
    - Will users enter correctly?
    - Will they remember them, share them, choose weak ones or write them down?
- Something you have works until the user doesn't have it
  - Need to think about the user and their life
  - We carry phones everywhere, but not CAP devices
- Something you are works, but users find it to be sensitive

See Bonneau et al for an overview.

Category	Scheme	Described in section	Reference	Usability: *Memorywise-Effortless*	*Scalable-for-Users*	*Nothing-to-Carry*	*Physically-Effortless*	*Easy-to-Learn*	*Efficient-to-Use*	*Infrequent-Errors*	*Easy-Recovery-from-Loss*	Deployability: *Accessible*	*Negligible-Cost-per-User*	*Server-Compatible*	*Browser-Compatible*	*Mature*	*Non-Proprietary*	Security: *Resilient-to-Physical-Observation*	*Resilient-to-Targeted-Impersonation*	*Resilient-to-Throttled-Guessing*	*Resilient-to-Unthrottled-Guessing*	*Resilient-to-Internal-Observation*	*Resilient-to-Leaks-from-Other-Verifiers*	*Resilient-to-Phishing*	*Resilient-to-Theft*	*No-Trusted-Third-Party*	*Requiring-Explicit-Consent*	*Unlinkable*
(Incumbent)	Web passwords	III	[13]			●		●	●	○	●	●	●	●	●	●	●		○							●	●	●	●
Password managers	Firefox	IV-A	[22]	○	●	●	○	●	●	●	●	●	●	●		●	●	○	○						●	●	●	●	●
	LastPass		[42]	○	●	●	○	●	●	●	●	●	●	●		●		○	○	○	○		○	●	●		●	●	
Proxy	URRSA	IV-B	[5]	●				●		○			●	●	●				○			○			●			●	
	Impostor		[23]	○	●	●		●			●	●	●	●	○		●	●	○				○		●			●	●
Federated	OpenID	IV-C	[27]	○	●	●	○	○	●	●	●	●	●		●	●	●	○	○	○	○		●		●		●		
	Microsoft Passport		[43]	○	●	●	○	●	●	●	●	●	●		●	●		○	○	○	○				●		●		
	Facebook Connect		[44]	○	●	●	○	●	●	●	●	●	●		●	●		○	○	○	○				●				
	BrowserID		[45]	○	●	●	○	●	●	●	●	●	●		○	○	●	○	○	○	○				●		●		
	OTP over email		[46]	○	●	●		●		●	●	●	●		●	●	●	○	○	○	○		●	●	●		●		
Graphical	PCCP	IV-D	[7]			●		●	○	○	●	●	●		●		●			●	○		●	●	●		●	●	●
	PassGo		[47]			●		●	○	○	●	●	●		●	○	●			●					●		●	●	●
Cognitive	GrIDsure (original)	IV-E	[30]			●		●	○	○	●	●	●		●		●		●								●	●	●
	Weinshall		[48]			●		●			●	●	●		●		●	○	●				●		●		●	●	●
	Hopper Blum		[49]			●		●			●	●	●		●		●	○	●				●		●		●	●	●
	Word Association		[50]			●		●	●	○	○	●	●		●		●										●	●	●
Paper tokens	OTPW	IV-F	[33]					●		○	●	●	●		●	●	●		●	●	●	●	●	●	●		●	●	●
	S/KEY		[32]	●				●		○	●	●	●		●	●	●		●	●	●	●	●	○	●		●	●	●
	PIN+TAN		[51]					●		○	○	●	●		●	●	●		●	●	●	●	●	●	●		●	●	●
Visual crypto	PassWindow		[52]	●								●	○		●	●		○	●	●	●	○	●	●	●		●	●	●
Hardware tokens	RSA SecurID	IV-G	[34]					●	○	○		●	●		●	●		●	●	●	●	●	●	●	●		●	●	●
	YubiKey		[53]					●	○	○		●	●		●	●		●	●	●	●	●	●	●	●		●	●	●
	IronKey		[54]	○	●		○	○	○	○		●		●	●	●		●	○			○	●	●	●		●	●	●
	CAP reader		[55]					●	○	○					●	●		●	●	●	●	●	●	●	●		●	●	●
	Pico		[8]	●	●		●		○	○						●		●	●	●	●	●	●	●	●		●	●	●
Phone-based	Phoolproof	IV-H	[36]			○		●	○	○		○	○	○			●	●	●	●	●	●	●	●	●		●	●	●
	Cronto		[56]			○		●	○	○		○	○		●	●		●	●	●	●	○	●	●	●		●	●	●
	MP-Auth		[6]			○		●	○		○	○	○				●		○					●	●		●	●	●
	OTP over SMS			●	●	○		●		○	○	○	○		●	●	●	●	●	●	●	○	●	○	●		●	●	●
	Google 2-Step		[57]			○		●	○	○	○	○	○		●	●		○	○	●	●		●	●	●		●	●	●
Biometric	Fingerprint	IV-I	[38]	●	●	●	○	●	○			○					○	●	●		●						●	●	
	Iris		[39]	●	●	●	○	●	○			○					○	●	●		●						●	○	
	Voice		[40]	●	●	●	○	●	○			○	○			○	○	●	○		●						●	●	
Recovery	Personal knowledge		[58]	○		●		●	○	○	●	●	●		●	●	●								●	●	●	●	●
	Preference-based		[59]	○		●		●	○	○	●	●	●		●	●	●								●	●	●	●	●
	Social re-auth.		[60]			●		●	●	●	●	●	●		●	●	○	○							●	●	●	●	○

●= offers the benefit; ○= almost offers the benefit; *no circle* = does not offer the benefit.
= better than passwords; = worse than passwords; *no background pattern* = no change.
We group related schemes into categories. For space reasons, in the present paper we describe at most one representati
scheme per category; the companion technical report [1] discusses all schemes listed.

Table I
COMPARATIVE EVALUATION OF THE VARIOUS SCHEMES WE EXAMINED

Warnings

- Habituation as a constant problem
  - Engineers expect warnings to be processed via conscious reasoning, but over time it switches to automated reactions
- Users click-through pop-up warnings at different rates and speeds
  - Click through often rapid – System 1/Skill-based
  - UI + content of warning is important
  - Chrome click-through rate (70%) far higher than Mozilla's (33%) bc Mozilla hid the ignore option
- Visualizing security is also hard!!
  - What does the green lock in the URL bar mean vs what do users think it means?

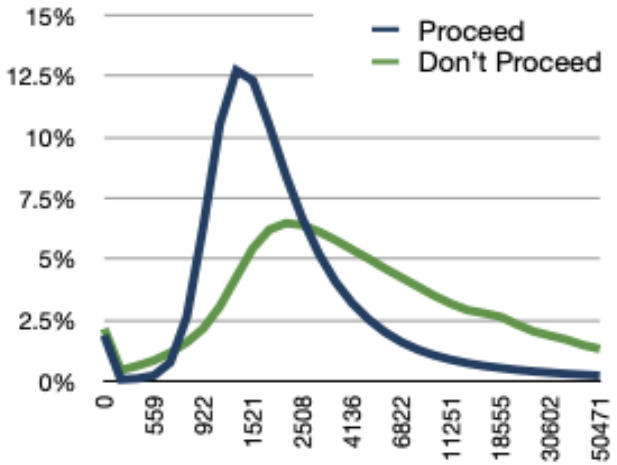
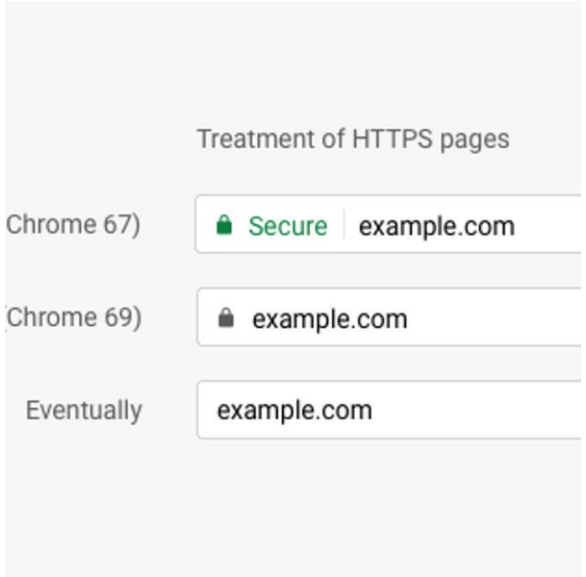


Figure 6: Google Chrome SSL clickthrough times (ms), by outcome. The graph shows the percent of warning impressions that fall in each timing bucket. The x-axis increases logarithmically, and we cut off the distribution at 90% due to the long tail.



Google says security should be so normal you don't see it. But bad security should be in your face.

Rules of thumb

- Think about defaults!
  - Microsoft disabled Macros by default
  - Build security into onboarding flow (e.g. MFA on by default)
- People will spend only so much time obeying rules – the compliance budget
  - Understand it and choose rules that matter
  - Requiring a 16-character password AND forcing a change every 90 days --> users write down password
  - Habituation to security warnings
- Rule violations are often an easier way of working, and sometimes necessary
  - Watch, measure and adapt
  - Password policy violations show authentication is too onerous

Google wants to enable multi-factor authentication by default

By Sergiu Gatlan

May 6, 2021 05:05 PM 0

"Soon we'll start automatically enrolling users in 2SV if their accounts are appropriately configured," This move is meant to increase Google user accounts' security by removing the "single biggest threat".

Summary of the lecture

- Understand the range of "human errors" that impact security
  - Contrast deep social engineering for romance/investment fraud against how phishing attacks unconscious processes, and how BEC attacks org rules
- Map these human errors to Rasmussen's SRK model
  - Processing based on knowledge vs rules vs skills
- Understand how threat actors increase the likelihood of errors
  - Time + pressure, authority, love, curiosity etc
- Be able to evaluate the efficacy of common mitigations and integrate these considerations into systems design
  - Training and warnings
  - Mental models and defaults