Week 2
Usability and Security

THE UNIVERSITY OF SYDNEY

Dr. Thilini Dahanayaka

School of Computer Science, The University of Sydney

Agenda

- Usability of security
- Biases in the human mind
- Decision making
- Deception in practice
- Passwords
- UX/UI design for security

Recommended Reading

- Security Engineering - 3rd Edition by Ross Anderson
  - **Chapter 3** – Psychology and Usability

1. Usability of Security

- “Only amateurs attack machines; professional target people”
- “People are the weakest link in the cybersecurity chain”
- “80%-90% of breaches are caused by human error”

Usability of Security

- The study of usability of security or usable security, is a very wide field.
  - Very fast developing
  - Surprisingly late recognised as a fundamental factor in security
- Many principles from usability translate directly to usable security.
  - Usability in software engineering refers to how easy and intuitive it is for users to interact with a software application (E.g. GUI design)
- Some things we will look at: are user psychology and mental models.

Problem: Theoretical vs. Effective Security

- Some mechanisms are very strong in theory, but fail when deployed in real-world environments.
- Security mechanisms must be usable and operationally sustainable, not just technically secure.
- **Example:** Smart cards vs. Passwords
  - Both provides clear mental model the user (Unlock a door using a key or Unlock using a secret)

Smart cards

- Store high-entropy cryptographic keys
- Hardware-backed protection
- Resistant to guessing attacks
- Clear “possession” mental model (e.g. door key)

However, multiple usability challenges

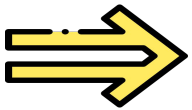
- But what if I forget them at home?
- What if the reader malfunctions?
- What if not all devices or OSes support them?
- What if I sit on them? Or leave them in the car on a sunny day?
- Some real pain points in using them!

Problem: Theoretical vs. Effective Security

- Some mechanisms are very strong in theory, but fail when deployed in real-world environments.
- Security mechanisms must be usable and operationally sustainable, not just technically secure.
- **Example:** Smart cards vs. Passwords
  - Both provides clear mental model the user (Unlock a door using a key or Unlock using a secret)

Passwords

- Cryptographically weaker
- No hardware dependency
- Low deployment friction
- Fewer ecosystem constraints



Users find "passwords" "easy to use"

- Immetrial
- Conceptual model matches user's intuition

Fundamental Security Design Principles (Last Week)

- Recall from the previous lecture
  - Economy of mechanism
  - Fail-safe defaults
  - Complete mediation
  - Open design
  - Separation of privilege
  - Least privilege
  - Least common mechanism
  - **Psychological acceptability**
  - Isolation
  - Encapsulation
  - Modularity
  - Layering
  - Least astonishment

Principle of Psychological Acceptability

- The original definition comes from J. Saltzer and M. Schroeder, "*The protection of information in computer systems*", Proc. IEEE 63:9, 1975.

The Principle of Psychological Acceptability says security mechanism should not make a resource more difficult to access than if the mechanism were not present (Bishop, 2003).

- **Example**: Consider setting up file permissions
  - Programmers find it easy to set file permissions correctly. A secretary may find it much harder.
- **Alternative Definition:** Security mechanisms cannot be applied in a vacuum. They must take into account abilities, knowledge, and mental models of people using it.
  - Often developers design the mechanism based on their own expectations of these people, which is far away from the reality. Next we discuss some examples to demonstrate this.

a. Acceptability of Passwords

- **Psychological acceptability** demands a low threshold for users to use the mechanism correctly - this can conflict with the goal of higher security.
  - **E.g.** Administrators understand the need for strong passwords - live in a world ‘under attack’.

- Users generally do not!
  - Use names, places, words from the dictionary as passwords (predictable)
  - Or different language - attacker has many dictionaries!

Top passwords in 2025

UNITED STATES			GLOBAL		
	PASSWORD	COUNT		PASSWORD	COUNT
1	admin	371,818	1	123456	21,627,656
2	password	263,126	2	admin	21,030,012
3	123456	257,124	3	12345678	8,274,408
4	12345678	115,041	4	123456789	5,673,712
5	123456789	75,156	5	12345	3,950,777
6	12345	57,435	6	password	3,545,119
7	Password	49,338	7	Aa123456	2,520,728
8	12345678910	45,673	8	1234567890	1,418,939
9	Gmail.12345	40,625	9	Pass@123	1,210,039
10	Password1	35,529	10	admin123	1,087,247

a. Acceptability of Passwords (contd.)

- Best-known mechanism (can change soon). Users understand them intuitively!
- Principle of psychological acceptability: passwords must add only minimal overhead
- Poor choices can be easily attacked:
  - Relate to user's environment? Easy - check social media!
  - Dictionary attack: Try words from the dictionary ('security') and variations following some substitution algorithm ('s3cu4itY')
  - Brute-force: try all character combinations
- **Unpredictability** helps against dictionary attacks and guesses
- **Length** helps against brute-force attacks
- **Problem:** unpredictability collides with the human capacity to memorize!

b. Acceptability - Other Examples

- Patching
  - A patch is an update to a system (often: software) to enhance functionality or remove a problem – often, a security problem.
  - Collecting all necessary patches
  - Conflicts in specific systems
  - Ideally, the process of maintaining system security should be invisible – this immediately collides with acceptability in this case.
  - Automatic updates?

- Security configurations
  - Just building a secure system does not yet provide security: it must be configured correctly for the environment it is going to operate it.
  - Do you use all the security/privacy features in your computer, or in social medial.

Categories of Human Errors

Manual Skill Failures

- Actions performed often become a matter of skill, but we can slip when a manual skill fails.
- **E.g.,** Typo squatters - users type the wrong URL accidentally.

Slips and lapses at the level of skill

Following the wrong rule

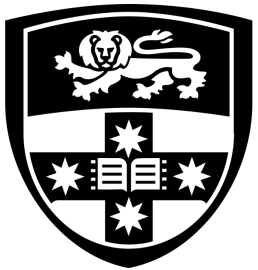
- Actions that people take by following rules are open to errors when they follow the wrong rule.
- **E.g.,** Trust HTTPS or the bank's name

Mistakes at the level of rules

Cognitive reasons

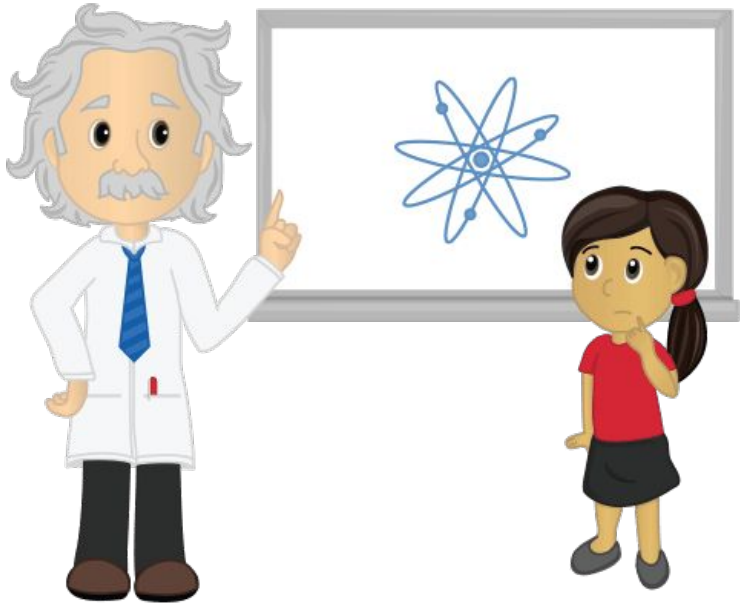
- People do not understand the problem or pretend they do.

Limitations at cognition level



THE UNIVERSITY OF
SYDNEY

2. Biases in the Human Mind



Source: https://www.allaboutlearningpress.com/blog/curse-of-knowledge/

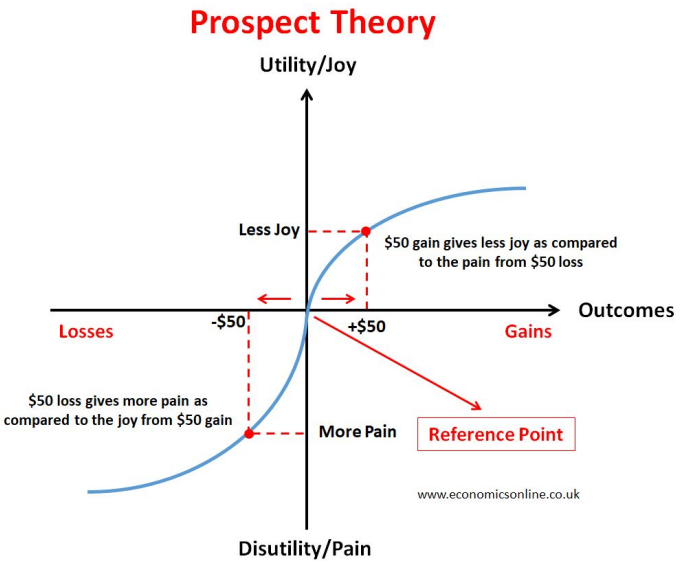
Prospect Theory and Risk Misperception

Prospect theory: People dislike losing and use all sorts of heuristics to make decisions. Successor of expected utility theory.

"Don't miss out, sale ends soon!" can be more persuasive than highlighting the gain "Get it now, On sale!"

"Your PayPal account has been frozen, and you need to click here to unlock it."

Risk Misperception: If we like an activity, we tend to judge its benefits to be high and its risk to be low. Conversely, if we dislike the activity, we judge it as low-benefit and high-risk.



Prospect Theory and Risk Misperception

Anchoring effect: We base a judgment on an initial guess and then adjust it if necessary

In lottery scams, the promise of a large jackpot serves as an anchor, making a small 'processing fee' seem negligible by comparison, leading victims to pay the fee in hopes of claiming the non-existent prize.

Availability heuristic: Tendency to think that whatever is easiest for us to recall should provide the best context for future predictions

'Spray and pray' traffic ticket SMS scams by referencing common and memorable traffic violations, prompting recipients to recall their own driving experiences and pay fines without verifying the ticket authenticity.

Present Bias and Hyperbolic Discounting

Present bias refers to the tendency of people to give stronger weight to payoffs that are closer to the present time when considering trade-offs between two future moments

- Causes people to decline security updates → vulnerable
- Privacy paradox → People say they care about privacy. But don't take time to
    - Check cookie notifications
    - Read privacy policies
    - Certificate errors

Clustering Illusion

Clustering Illusion arises from the human tendency to seek order and structure, even when none exists, leading individuals to infer relationships, regularities, or causes in purely random variation. This can reinforce irrational beliefs about hidden patterns or inconsistencies and may result in flawed reasoning and poor decision-making.

- Famous story: ‘hot hands’ belief
  - ‘Basketball players will score again if they just scored’.
  - Data shows absolutely no evidence for this.
  - Spectators and coaches believed the data was wrong

- Clustering illusion can be exploited. One such example is in phishing attacks, where scammers send out emails that appear to come from legitimate sources, such as a bank or a social media platform, and try to convince the recipient to click on a link or provide sensitive information.
- To make the phishing email appear more legitimate, the scammers may use personal information or data that they have obtained about the victim.
- They may also create fake logos, graphics, or URLs that resemble those of the legitimate organization.

Confirmation Bias

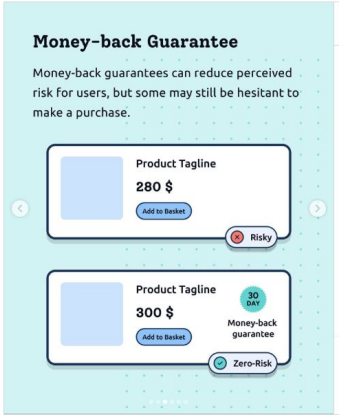
Confirmation bias is the tendency to seek, interpret, and remember information in ways that confirm one's existing beliefs or hypotheses, while ignoring or discounting evidence that contradicts them.

- Humans find it easier to accept something as true if it conforms to their own preconceptions.
- Inversely, it can be very hard to convince them that something is wrong.
- Once the belief is established, users will find it easier to ignore an inconsistency in data that they are presented with.
- This can be exploited, especially in Social Engineering

Zero-Risk Bias

Zero-risk bias is the tendency to prefer options that completely eliminate a specific risk (achieving 0%), driven by a desire for absolute certainty, even when alternative options would reduce overall risk by a larger amount.

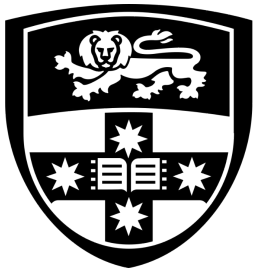
- Common in government, crypto, and sometimes in technical communities.
- Preference for absolute, information-theoretic security.
- Highly secure but imperfect solutions are dismissed.
- Eliminating one risk (0%) is preferred over reducing total risk more.
- **Classic study:** A study asked participants to rank cleanup approaches for hazard sites
    - Approaches that cleaned one site perfectly, but did poorly for others, were preferred to those that reduced **overall risk** to a minimum.



Source: https://www.instagram.com/p/CqlBN9fK89G/?img_index=3

Biases and the Importance of Defaults and Nudges

- Human decision-making is systematically biased.
- Because of these biases, **defaults** matter - people often choose the easiest option (present bias).
  - **Defaults** are pre-selected system settings applied unless users actively change them.
- Giving too many configurations to the users can be counter-productive
  - **Control paradox:** providing the illusion of control causes people to share more information.
- **Nudges** are subtle, non-coercive prompts or design features that guide users toward better decisions.
  - They preserve choice.
  - They encourage desired security behaviours (e.g., updates, 2FA, privacy review).
- Effective security design combines good defaults + well-designed nudges.



THE UNIVERSITY OF
SYDNEY

3. Decision Making

a. Classes of Techniques to Influence People

These are listed separately because they come from a different origin than psychology – Within the scope of this class you can consider these as cognitive biases as well.

1) **Reciprocity:** people feel the need to return favors
   - Can be exploited, e.g.,
     - Helpful caller from the IT department walks the new employee through software
     - Then asks them to install that “new tool that IT is rolling out”
     - Even though installing new software may be against company policy, there is a chance that the employee will install the tool (malware)
   - **Also called as reciprocation bias**

Classes of Techniques to Influence People

2) Commitment and consistency: People suffer cognitive dissonance if they feel they are being inconsistent.

- Can be exploited, e.g.,
  - Attacker calls new employee and advises them of security policies, to which the employee makes an explicit commitment
  - Then asks employee for password to ‘verify their compliance’
  - Doesn’t stop here
  - Once password is revealed, attacker asks user to choose a ‘better’ one - one that the attacker can easily guess

Classes of Techniques to Influence People

3) Social proof/validation: People want the approval of others. Most people tend to comply with something when they see others do it as well.

- Can be exploited, e.g.,
  - Caller says they are conducting a survey and gives names of people in the same department who already did it
  - Victim takes part in the survey and reveals sensitive data

4) Like bias: People tend to comply to requests coming from people they like.

- Can be exploited, e.g.,
  - The attacker may claim to have similar interests, beliefs, or attitudes—or they may make compliments
  - Most humans have a natural desire to be liked by others as well, and most want to be seen as helpful

Classes of Techniques to Influence People

5) Respect to authority: People are deferential to authority figures

- Experiment (R. Caldini, Influence, 2006)
  - Nurses in the hospital received a call asking them to administer medication to a patient—the caller claimed to be a hospital physician
  - In 95% of cases, nurses complied (but the experiment was broken off before they could go to the patient)
  - Similarly an attacker can ask for a bank transfer

6) Scarcity: We are afraid of missing out

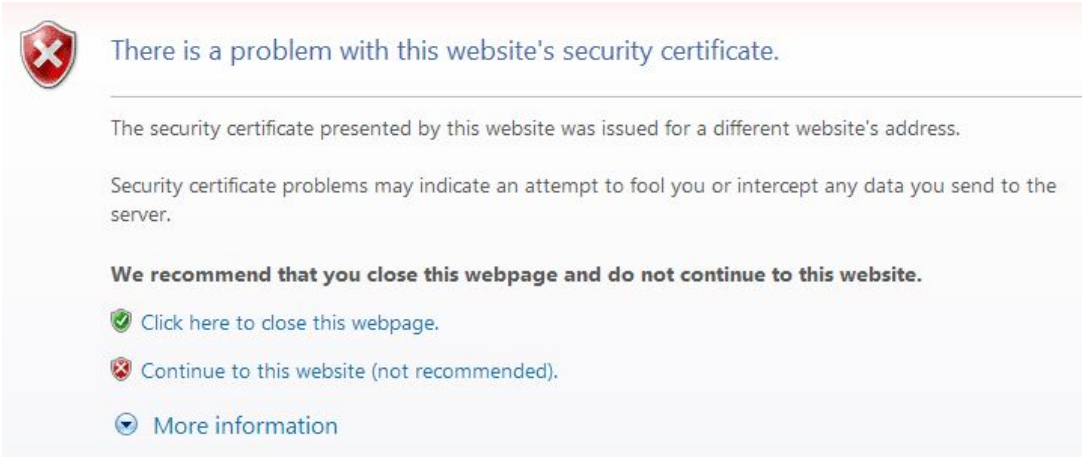
- The more difficult it is to acquire an item the more we value it. In other words, if we think something is scarce, we want it more.

b. User Conditioning

- Decision making is also affected by automatisms introduced by conditioning
- User conditioning refers to getting users habituated to react to certain situations in a specific way (‘click-whirr reaction’)
- Effect: user does not stop to think when a security decision has to be made but reacts automatically
- Example: warning dialogues
  - Nagging dialogues that are unnecessary and need to be ‘cancelled’
  - ‘Click-away’ response without reading the warning!
- Example you may all know: how often did you delete a file accidentally?
  - And after a warning?

b. User Conditioning

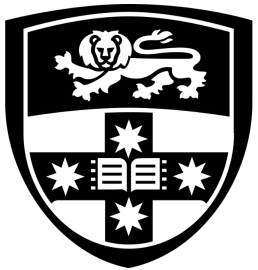
- **Example:** Responding to certificate error
  - Most errors benign - attack in practically 0% of cases
  - Users completely right to ignore
  - So, what to report to the user to avoid conditioning



c. User Education

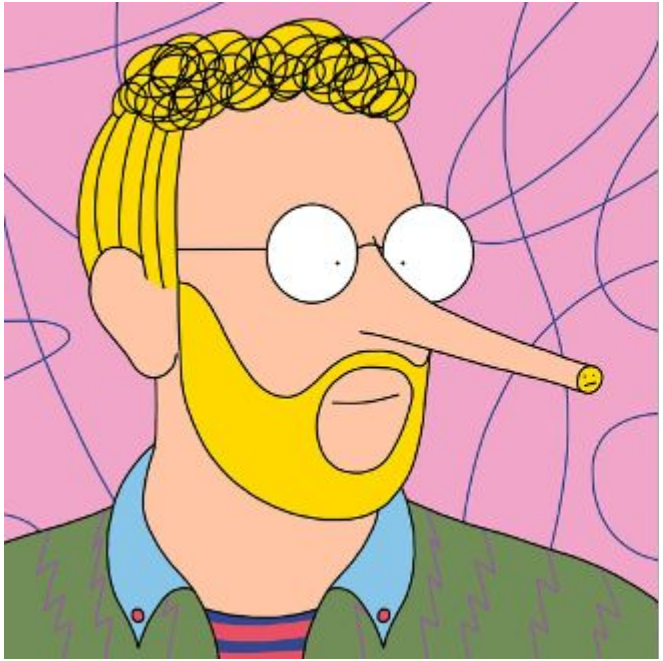
- Attempts to educate users have met with mixed success
  - Just telling them (e.g., distributing company policy) usually does not have an effect
  - Training, i.e., incorporating feedback and iterations, provides better results
  - This seems natural, given our understanding of biases and workings of the human mind
- Complication: Security language is fundamentally different from user language
  - Significant cost for the user
  - Users have no intrinsic motivation to learn security - **they have jobs to get done**

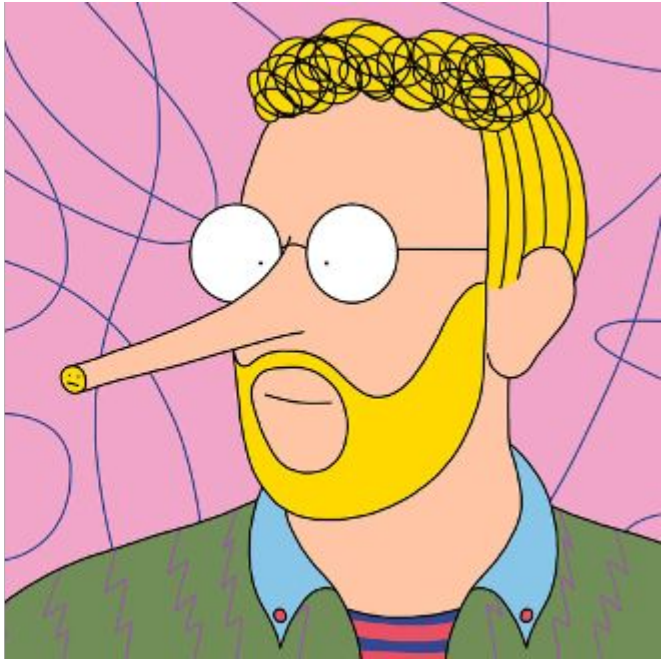
User education improves security, but not as much as we often assume; human behavior is variable and easily influenced by context, pressure, and sophisticated attacks. Therefore, training should complement, not replace, strong technical and systemic security controls.



THE UNIVERSITY OF
SYDNEY

4. Deception in Practice





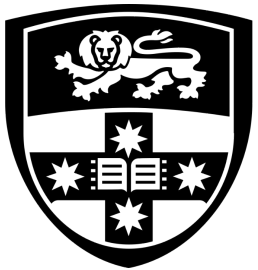
Social Engineering and Phishing

Social Engineering

- Hacking systems through the people
  - Old scam from the 20th century is to steal someone's ATM card and then phone them up pretending to be from the bank
  - As a result of an audit test in 2007, 62 out of 102 IRS employees at all levels provided their IDs and changed their passwords to a known value
- Very useful for stealing credentials and identities

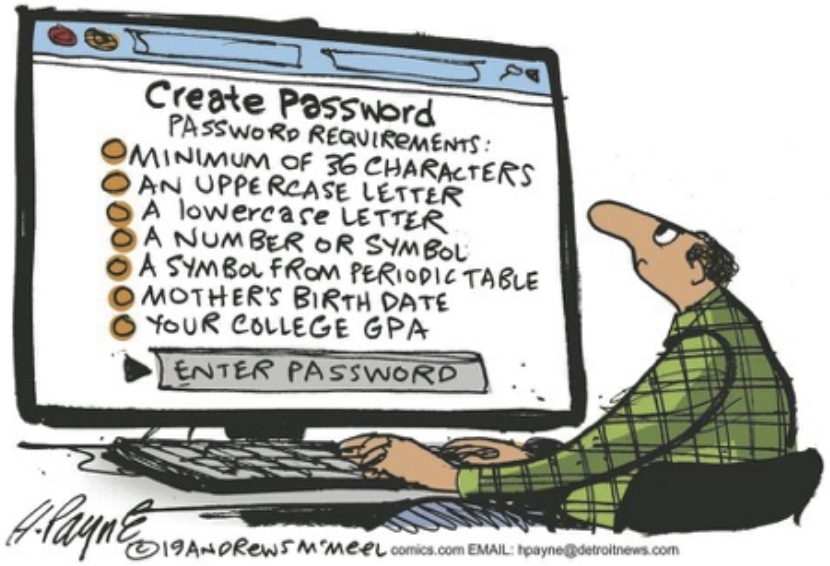
Phishing

- Phone-based social engineering was the favoured tactic of the 20th century. Now it is phishing.
  - In 1996, 336 computer science students at the University of Sydney were sent an email asking for their password to 'validate' the password database after a suspected break-in. 138 returned a valid password, 30 returned an invalid password, and 200 changed their passwords. But every few of them report to the authority.
  - Attackers often reused genuine bank emails with URLs changed or fake login windows (www.paypals.com/login)



THE UNIVERSITY OF
SYDNEY

5. Usability and Security of Passwords



Three Broad Concerns

- Will the user enter the passwords correctly with a high enough probability?
- Will the user remember the password, or will they have to either write it down or choose one that’s easy for the attacker to guess?
- Will the user break the system security by disclosing the password to a third party, whether accidentally, on purpose, or as a result of deception?

Concern 1 - Password Entry

- Users might have difficulty entering the password correctly if the password is too long or complex.

- Since 2010, password rules were ‘at least 1 lowercase letter, uppercase, number, special character’ → annoying and fiddly

- Users like to use a simple password such as passphrases of three or four words to enter it without making many errors

Concern 2 - Remembering the Password

- In fact, standard password advice can be summed up as:
  - “Choose a password you can’t remember, and don’t write it down”

- Is writing passwords down really a bad idea?
  - May be not: That depends on the context
  - In an open office environment/public place – Definitely No

Concern 3 - Disclosing the Password

- Password entry:
  - Interface design: The design of EFTPOS, for example, some customers don’t cover when entering their PIN in public
  - Technical defeats of password retry counters: Attackers might use the timing attack

- Password storage:
  - Password cracking: Brute-force attack, dictionary attack, various password cracking tools
  - Password manager: Tool to keep all your passwords stored securely but can we trust them? Apply Keychain, Chrome/Mozilla/Edge/Safari built-in managers

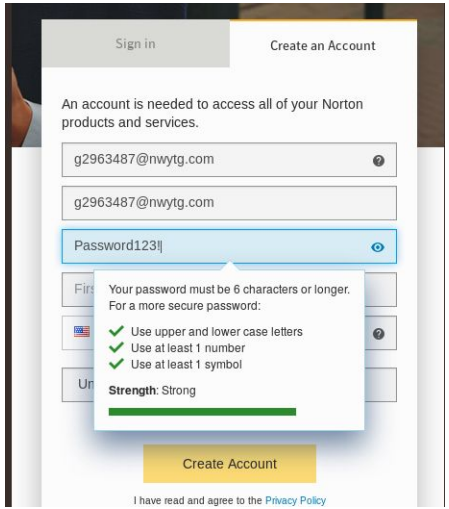
Outdated Advice

- ‘Passwords must be changed every 30/60/90/180 days’
  - In general, this is **bad** advice
  - Stored correctly, **strong** passwords remain **strong**
  - Humans can remember a handful of strong passwords at best
  - When forced, they will choose ever-weaker passwords
  - Passwords for different logins become more similar over time
  - People run out of ideas for memorable passwords

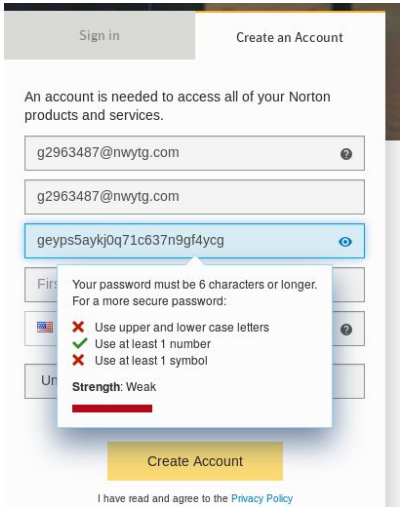
- **Do not expire passwords unless you suspect they have been compromised**.
  - UK policies have recommended this since 2015.
  - Australia followed in 2017.

Outdated Advice - Complexity Rules

‘Your password must consist of 6-12 characters, with at least one special character and two numbers’ Why is this bad? Let’s have an example.



Password123! Is accepted



A good password is rejected

Complexity Rules != Complexity Rules

- Passwords still need to be complex to be unpredictable!
- So how do we test for complexity?
- Bad: compliance with a complexity rule!
- **Much better—and cheap checks:**
  - Check if the password is from a dictionary (in any language)
  - Check if the password is on the list of passwords tried by attackers
  - Check for common substitutions as well: 4 for A or R, 3 for e, 7 for L etc. Check for combinations of numbers that are really dates
- **Rationale:**
  - Attackers try dictionaries first—much more cost-efficient!
  - Brute-force is a last resort!

Tension: Unpredictable vs. Memorable

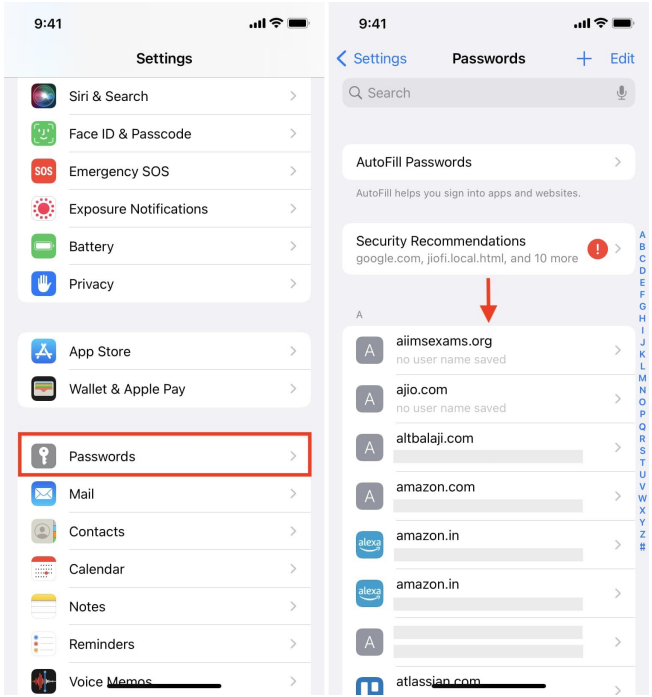
- Passwords must still be **unpredictable** and **memorable**
  - Implies complexity and length
  - Tension to memorability

- How can we bridge that gap?

- The following methods are tried and tested:
  - Password managers
  - Passphrases via Dice method (or similar)
  - ‘First letters from a sentence’ method (but careful)
  - PAO method (but careful)

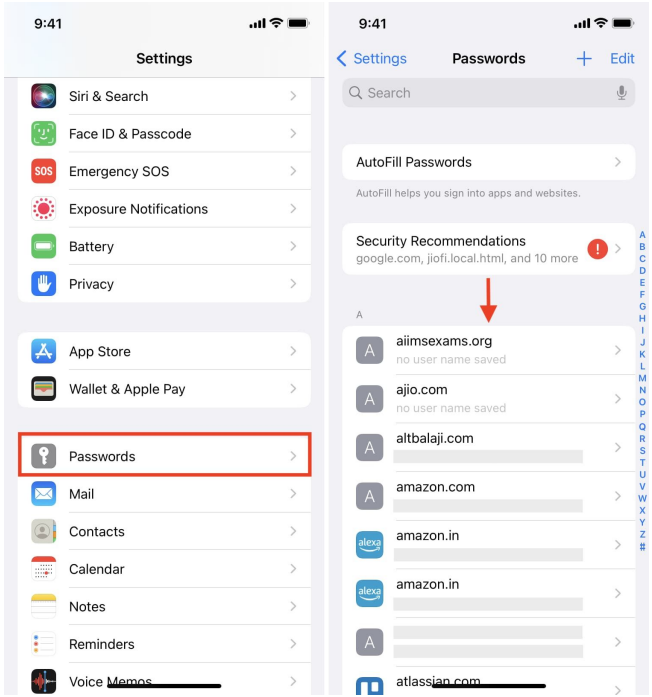
Password Managers

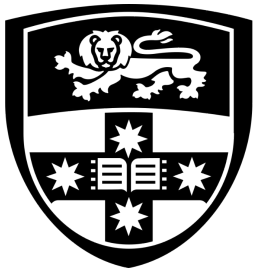
- **Concepts:**
  - Strong master password protects database of passwords and logins
  - Can generate highly complex and long passwords for you
  - Often integrate with browsers and other applications
  - Can copy & paste into the application
  - Some are also available for mobile OSes



Password Managers - Drawbacks

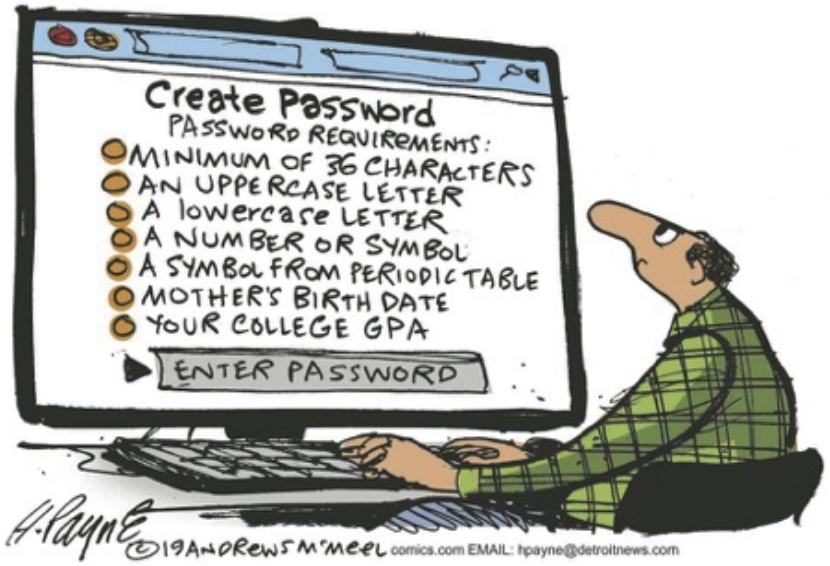
- Loss of master password
  - Means loss of all passwords
- Portability remains an issue
  - Not widespread on mobile devices
  - Things are better now
- Longer workflow (again getting better)
- Some sites prevent copy and paste
  - They are wrong!





THE UNIVERSITY OF
SYDNEY

6. Creating Secure and Memorable Passwords



Passphrase and Dice Method

- **Concept**: Randomly choose 7-12 words from the dictionary
  - Words are easier to memorize
  - In particular, if we can link them with a ‘story’
  - Easily extendable: substitute words with characters etc.

- Dice method: standard dictionary, dices for random picks
- Provides reasonably high security!
- Drawback: much typing

Letters-from-a-sentence

- **Concept**: choose a memorable sentence; pick 1-2 (first) letters to create a password

- Example: Wow! 62 students, all in this memorable class of 2026!
  - Ideally, has upper and lower case words
  - Ideally, has digits (e.g. dates, amounts)
  - Ideally, has special characters

- Easily extended: W!62s,aitmc02o26!

PAO (Person-Action-Object) Method

- **Concept:**
  - Uses visual cues and unusual imagery to jog memory
  - Choose images of place and person
  - Imagine random action on random object involving this place and person
    - E.g. Darth Vader riding a pony on Mt Everest
  - Have some rule to form password from that
    - E.g. DVrapoMtE!
    - Easily extended: DVr4poMt3!

Password Practicalities

- In theory, we should **use different passwords for every site**
  - It is not transparent how sites protect your password
  - Compromise of one site can yield username (often: email) and password
- In practice, most of us have dozen of accounts (if not more).
  - Not possible to memorize them all.
- Password managers are a great choice if:
  - You don’t need all passwords with you all the time (mobility)
  - You understand the ideas behind them well enough

Password Practicalities

- Again, what about **writing passwords down**?
  - What is the risk of someone evil getting the password vs. you forgetting it and hence choosing an easy one?
  - Writing down can be OK at times, e.g., for websites if you only log in from home, WiFi passwords, etc.
  - Unless your family is your attacker or you have many unknown visitors
  - Normally **not suitable** for office situations, admin duties etc.

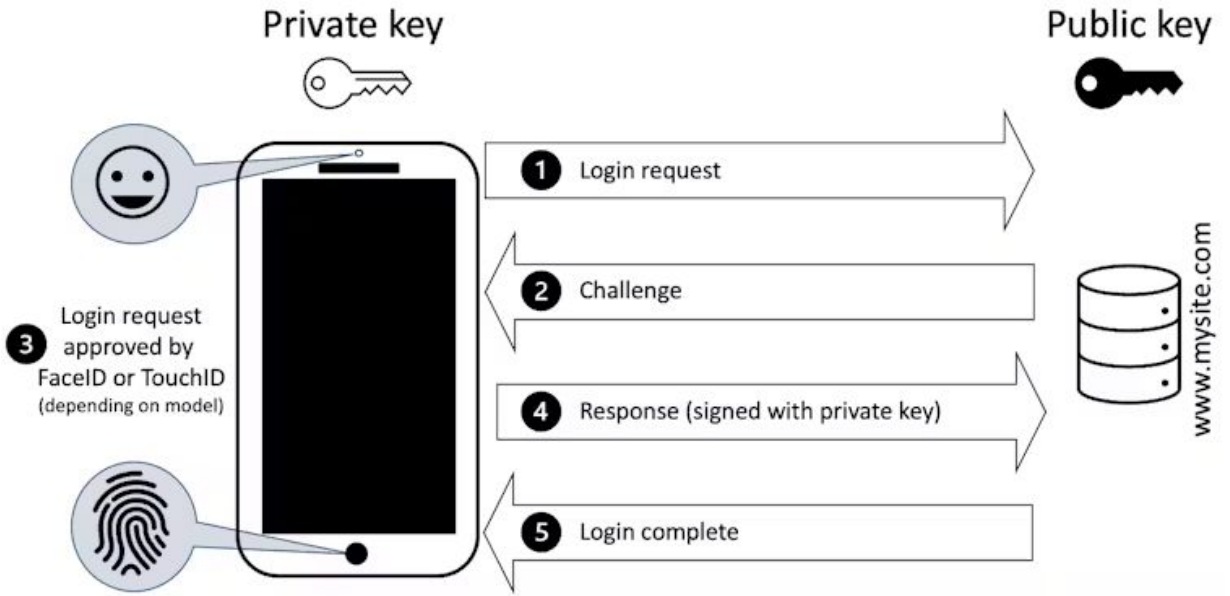
- **Hybrid approaches** may be the best solution
  - Password managers where they work for you
  - A handful of strong passwords that you use for sites that must always be accessible

Will Passwords Eever be Outdated?

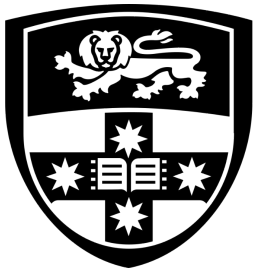
- Use SMS to authenticate?
- Authentication token?
- Biometrics?

Example: Apple, Microsoft,Google and many others are promoting “passkeys”

Biometric Passkeys



https://www.smartcompany.com.au/technology/apples-passkeys-update-passwords/

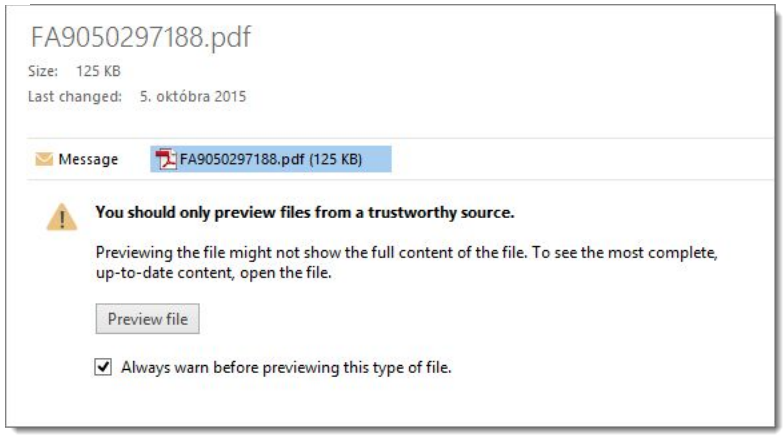


THE UNIVERSITY OF
SYDNEY

7. UX/UI Designs for Security

Figure source: Fixing bad UX designs/Packt

UX/UI Designs for Security



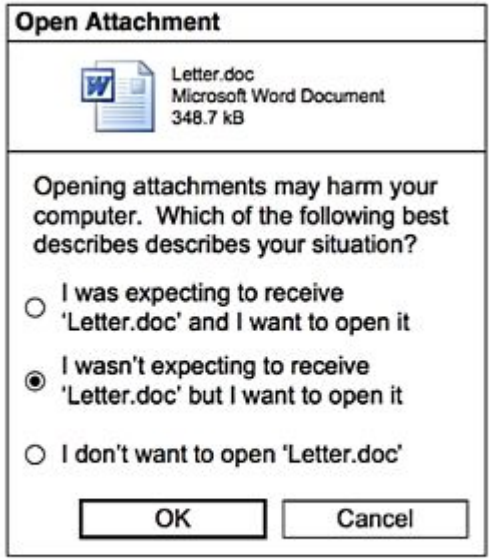


Figure: Source: Gutmann, 2014.

Mental Models

- Mental models describe how users perceive a system
  - Borrows experience from related situations in users' lives
  - When known, designers can use them to steer users towards secure decisions
- Several companies have become very proficient in exploiting mental models
  - Apple, Google, …
- Several factors at play that :
  - **Affordance** - elements that guide a user towards correct use
  - **Constraints** - user cannot easily make a wrong choice
  - **Conventions** - recognisable elements (e.g., pause symbol on DVR, ...)

Mental Models

Affordance: Perceiving a button on a screen, we know it is something that can be pressed to produce an action.

Constraints: Users can still move around and interact but are confined to a certain area under certain parameters.

Conventions: Users can recognise the elements and their functionalities.

Take Away Messages

- If your security feature did not work, do not blame the user.
- Do not make the mistake of assuming users have developers' experience and hence experience a lighter amount of stress when making decisions.
- Design of a new security feature must take psychological factors into account.
- Use of design process that integrates stakeholders' perspective is recommended.
- Relying on user education **instead of the above** is not recommended.

Recap

- **We discussed:**
  - Principle of Psychological Acceptability
  - Biases in human mind
  - Influencing decision making
  - User conditioning
  - User education
  - Challenges in passwords
  - Biometric Passkeys
  - Ideas from usability to design security features