The University of Sydney
School of Computer Science
Dr Suranga Seneviratne
Senior Lecturer - Security

THE UNIVERSITY OF SYDNEY

Psychology and Usability

Recommended Reading

Security Engineering - 3^{rd} Edition by Ross Anderson

- **Chapter 3:** Usability and Psychology

These lecture notes are given to you to assist with understanding the lecture content better. This content is prepared based on the above book chapters. You are not allowed to upload this material to any internet source or share it with anyone else.

1 Usability of Security

People are the weakest link in the cybersecurity chain, and 80%-90% of security breaches are caused by human error. For example, during the Twitter bitcoin scam in 2020, the attacker breached the system using credentials for an administrator account, which were available on a Twitter Slack channel [1]. A main contributing factor to such vulnerabilities is that users behave in ways engineers/developers do not expect, as the users may not be as tech savvy as the developers, and their main focus will be on the tasks assigned to their role rather than the security of the system. Hence, the system developers need to find solutions that seamlessly help the users and ensure the security of the system. The field that explores such solutions is referred to as "*Usability of Security*".

The study of *usability of security* or *usable security* is a very wide field that is fast developing. This field is intricately connected to *usability engineering* or *usability* in the domain of human-computer interaction. For instance, in software engineering, usability denotes the degree of ease and intuitiveness with which users can interact with a software application. Similarly, the designers should focus on making cybersecurity applications more user-friendly.

2 Theoretical vs. Effective Security

For a long time, the security community focused on creating theoretically secure solutions that were often not user-friendly. There have been many examples of technologies that promise high security yet have been found to be unacceptable to users.

One such example is the use of smart cards. While smart cards offer a higher level of security compared to traditional passwords, their requirement for additional hardware and the complexity of their implementation have often led to resistance from users, demonstrating the need for balancing security with usability.

3 Principle of Psychological Acceptability

The principle of Psychological Acceptability is first defined by J. Saltzer and M. Schroeder [**?**].

Principle of Psychological Acceptability

It is essential that the human interface be designed for ease of use, so that users routinely and automatically apply the protection mechanisms correctly. Also, to the extent that the user's mental image of his protection goals matches the mechanisms he must use, mistakes will be minimized. If they must translate their image of their protection needs into a radically different specification language, they will make errors.

This definition can be summarized as **"a security mechanism should not make a resource more difficult to access than if the mechanism were not present"** [**?**].

Implementing this principle isn't always straightforward, as it requires consideration of the abilities, knowledge, and mental models of the individuals who will be using it. Unfortunately, developers often design mechanisms based on their own expectations of users, which can lead to issues. For example, programmers may find it relatively easy to configure file permissions correctly, while a secretary might struggle due to a lack of understanding about security. The following examples illustrate this further.

- **Acceptability of passwords**
  Administrators understand the need for strong passwords as they live in a world 'under attack'. However, users often fail to comprehend the requirement of strong passwords even though they understand passwords intuitively. Hence users resort to easily guessable passwords containing names, locations, or common dictionary words. Such poor password choices can be easily cracked, using one or a combination of following methods.
  - If it relates to the user's environment check their social media (social engineering)
  - Dictionary attack: try words from the dictionary ('`security`') and variations following some algorithm ('`s3cu4itY`')
  - Brute-force: try all character combinations

  A strong password should be unpredictable and long to prevent these attacks. Yet, unpredictability often clashes with the human capacity to memorise and contradicts the principle of psychological acceptability, which advocates for passwords that impose minimal overhead.
- **Acceptability of patching**
  A patch is an update to a system, typically software, designed to improve functionality or resolve issues, commonly security vulnerabilities. However, despite their critical nature, patching processes often present complications for users.

  e.g.:
  - Disruptions: Patching often requires users to temporarily halt their work or exit applications while updates are installed and may also necessitate system restarts to take effect fully.
  - Complexity: Some patching procedures may be complex or require technical knowledge, which can be challenging for less experienced users to navigate.

- Notification Fatigue: Users may become overwhelmed by frequent patch notifications, especially if they are inundated with alerts for minor updates or non-critical patches.
- Compatibility Issues: In some cases, patches may introduce compatibility issues with existing software or configurations, resulting in unexpected errors or malfunctions.

In an ideal scenario, maintaining system security should occur seamlessly in the background without disrupting user experience. However, the inherent complexities of patching often clash with acceptability. As a result, some companies have streamlined their patching processes to minimize user inconvenience. For example, with iPhones, when the device is plugged in at night, a notification may appear stating, "An update is available. It will be installed tonight while your device is charging and connected to Wi-Fi." This approach ensures user-friendliness by avoiding disruptions to the user's activities and ensures that the update is completed without draining the device's battery or consuming mobile data.

- **Acceptability of security configurations**
  Building an effective security system is only the beginning; true security is achieved when it's properly configured for its specific operating environment. While developers possess comprehensive knowledge of available configurations, users may lack the same expertise, often finding these settings intricate and time-consuming, which contradicts the principle of psychological acceptability. Failure to configure a system properly can lead to significant security vulnerabilities in the system.

  e.g.:
  - Many individuals overlook configuring security and privacy features on their social media accounts, leading to the inadvertent sharing of personal information.
  - In 2009, insurgents in Iraq gained unauthorized access to video footage captured by US drones. Subsequent investigations uncovered that the reason for the breach was due to lack of encryption in the communication channels [**?**]. It turns out that even if the encryption technology was available for those drones, they were not updated/enabled.

  An effective approach to enhancing the usability of security configurations involves system developers establishing default settings that provide the desired levels of security.

3.1 Categories of Human Errors

To construct security solutions resilient to human weaknesses, it's crucial to understand how users' mental models of systems work, how they differ from developers' mental models, and techniques that can be leveraged to explore users' learning processes and comprehension of systems.

The **safety research community** has dedicated significant efforts to analyzing human errors in equipment operations. These predictable forms of human error stem from the inherent nature of cognition. While our mental models excel in recognizing individuals, sounds, and concepts compared to computers, they also render us susceptible when an erroneous model is triggered. Human errors made while operating equipment fall into three broad categories.

These early ideas have been used by attackers to craft attack strategies and, as we will learn later, by the security engineering community to design secure solutions.

1. **Manual skill fails:** Actions performed often become a matter of skill; however, a failure in manual skill may lead to accidents (safety engineering) and vulnerabilities (security engineering).

   Examples:

   (a) In an industrial setting, a worker operating heavy machinery may become adept at performing routine tasks. However, if the worker becomes distracted or fatigued, their manual skills may fail, potentially causing a serious accident.

   (b) Typing errors with URLs can be exploited by typo squatters who register domain names similar to popular ones. Some examples are provided in Table 1.

Targeted Domain	Typosquat Domain Example
www.github.com	www.glthub.com (Typo)
www.google.com	www.gougle.com (Typo)
www.amazon.com	www.amozon.com (Typo)
www.netflix.com	www.netfllix.com (Duplicate "l")
www.linkedin.com	www.linkedin.co (Missing letter)

Table 1: Examples of Typosquatting

2. **Following the wrong rule:** Actions that people take by following rules are open to errors when they follow the wrong rule.

   Examples:

   (a) In the context of industrial safety, the rule "Always assume that a machine is safe to operate if the safety light is green" can be dangerous. There may be situations where the safety light system is malfunctioning or has not been updated to reflect a new hazard.

   (b) The rule "Always trust HTTPS because it is secure" is flawed, as attackers can also acquire certificates, undermining the reliability of HTTPS.

3. **Cognitive reasons:** Individuals may simply not understand the problem, or pretend that they do, or ignore the advice to get their work done.

   Examples:

   (a) Despite being trained on the new protocols, the operator finds them cumbersome and believes their old methods are sufficient. They choose to ignore the advice, leading to unsafe practices and potential hazards.

   (b) Despite being advised on the necessity of keeping their software up-to-date, the employee finds the frequent updates disruptive to their workflow. Believing that their system is secure enough, they choose to postpone or ignore updates, thereby increasing the risk of a security breach.

These ideas on sources of errors from safety have been used with early success in security engineering. For example, *affordance* is an idea we will discuss later in this class. However, to design more usable security solutions, we need to understand further these cognitive reasons for errors and the human decision-making process in general.

4 Biases in the Human Mind

There are multiple theories that explain the human decision-making process. The early idea of ***Rational Choice Theory*** explains how individuals make decisions by selecting the option that maximizes their utility or satisfaction. ***Expected Utility Theory*** is a more specific model within the broader Rational Choice framework. It deals with decisions under uncertainty and risk, where outcomes are not deterministic.

Rational Choice Theory and Expected Utility Theory Example

Rational Choice Theory Example: A consumer deciding which brand of cereal to buy in a supermarket. They consider factors like price, taste, and nutritional value to maximize their satisfaction (utility).

Expected Utility Theory Example An investor choosing between different financial investments. They consider the potential returns (utilities) and the risks (probabilities of different returns) to select the investment that provides the highest expected utility.

However, our decision-making is often neither rational nor perfect. There are many biases in our psyche and some psychological factors that affect our decision-making process. For example, according to rational choice theory, people should always choose healthy food from supermarkets. However, that always doesn't happen. Therefore, we need other theories to explain our decision-making process. To explain decision human decision-making related to security engineering, we resort to ***Prospect Theory and Risk Mispercpetion*** **and** ***Behavioural Economics.***

Prospect Theory and Risk Misperception

Prospect Theory and Risk Misperception: Kahneman and Tversky challenged the traditional Expected Utility Theory by showing that people do not always act rationally when making decisions under risk.

- **Loss Aversion:** The theory introduced the concept of loss aversion, which states that losses are felt more intensely than equivalent gains.
- **Framing Effects:** Decisions are influenced by how choices are presented or framed.

Behavioural Economics: Behavioral decision theory incorporates insights from psychology to explain why and how people deviate from rational decision-making. It examines cognitive biases, emotions, and heuristics, providing a more realistic description of human behaviour.

4.1 Prospect Theory and Risk Misperception

Prospect theory models the risk appetite in the decision-making process. According to the theory, individuals are inclined to prioritise avoiding losses over making rational choices based on probability and the expected utility of an outcome. For instance, as illustrated in Figure 1, the psychological value (i.e., say sadness) of losing $100 is greater than the psychological value (i.e., say happiness) of winning $100. As a result, framing an action as avoiding a loss rather than gaining a profit can make people more likely to take it.

- "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." has a high probability of getting user reaction.

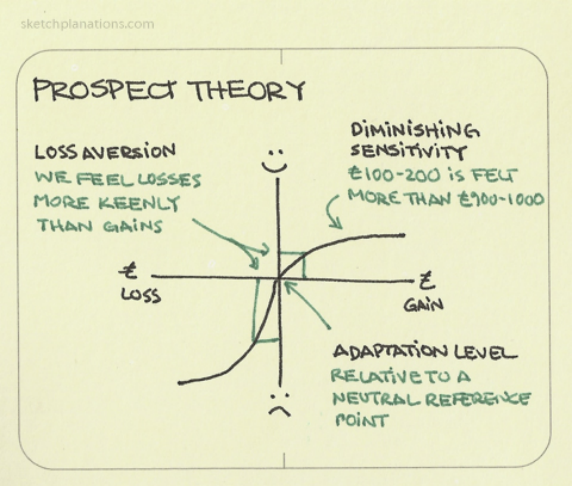


Figure 1: Prospect theory [?]

Moreover, people, in general, are bad at calculating probabilities and, hence, rely on various heuristics to aid in decision-making.

4.1.1 Risk misperception

When individuals express a preference for a particular activity, there is a tendency to perceive its benefits as substantial and its associated risks as minimal. Conversely, if an aversion exists towards the activity, individuals are inclined to evaluate it as having low benefits and high risks.

- A person who has a fear of flying, would perceive flying as riskier than driving, despite statistical evidence showing that flying is a much safer mode of transportation. This individual might opt for a long and arduous road trip instead of taking a short flight, driven by a fear that exceeds the objective risks associated with each mode of travel.

4.1.2 Anchoring effect

People tend to base a judgment on an initial guess and then adjust if necessary.

- When a person visits a car sale and sees a car advertised as $30K, they might think it too expensive and will be reluctant to buy it. But if the person is first told that the car was originally priced at $60K and has been discounted to $30K, the person will most likely anchor on the value of the discount and will be tempted to buy the car.
- In pop-ups promoting fake antivirus software, the scammers usually emphasize a substantially higher price than the original cost and then include a price reduction ('It is $1000 per year, we are giving it away for $500'). Victims are prone to anchoring their perception on the initial high price, thereby succumbing to the temptation to click on the provided link.

- 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. Same goes with Facebook Marketplace PayID scams [**?**].

4.1.3 Availability Heuristic

People tend to believe that information readily recalled, owing to its ease of accessibility, should serve as the optimal basis for future predictions.

- After viewing a documentary on shark attacks the previous night, an individual may exhibit hesitancy in visiting the beach the following morning, as vivid recollections of the specific details concerning shark attacks remain firmly embedded in their memory.
- 'Spray and pray' SMS scams at tax return time pretending to be from the ATO, prompting recipients to click links recalling their that they recently submitted their tax returns.

4.2 Behavioural Economics

Behavioral economics integrates insights from psychology with economic theory to better understand how individuals actually make decisions, as opposed to how they would if they were perfectly rational.

4.2.1 Present Bias and Hyperbolic Discounting

Present bias is the tendency of individuals to give more significance to immediate rewards or benefits over those that may be more substantial but are delayed in time.

- Users often postpone or ignore software updates and security patches as they find the immediate inconvenience of restarting or updating software more bothersome than the potential long-term consequences of leaving vulnerabilities unaddressed. They tend to give more weight to being able to complete their current task without interruption as more important.
- Users might swiftly click "Accept" on cookie pop-ups to remove the overlay and proceed to the desired content, rather than taking the time to review and adjust privacy settings as they perceive the immediate goal of accessing the website as more important than potential long-term consequences associated with the collection and use of personal data.
- Users might neglect reading privacy policies and adjusting privacy settings on online platforms or social media networks, prioritizing immediate access to features over long-term concerns about personal data exposure and privacy risks.

Hyperbolic discounting is a model used by decision scientists to quantify present bias. This means people use utility functions that discount the future so deeply that immediate gratification seems to be the best course of action, even when it isn't. We can explain explain the *privacy paradox* using the present bias and the hyperbolic discount.

Privacy paradox refers to the observed discrepancy between individuals' expressed concerns about privacy and their actual behaviors related to information sharing in the digital realm. Despite expressing privacy concerns, people often engage in behaviors that involve sharing personal information online such as ignoring security warnings, cookie notifications, and privacy policies.

4.2.2 Clustering Illusion

The clustering illusion refers to the tendency of individuals to perceive patterns or clusters in random or unrelated data. The clustering illusion arises because our brains are inherently wired to identify patterns, a trait that has historically provided an evolutionary advantage in various scenarios. Nevertheless, this innate capability can lead us astray when confronted with genuinely random data.

- The 'hot hands' belief in basketball where people believe that a player who has recently experienced success in making shots is more likely to continue making successful shots, despite having no statistical evidence to prove this idea.
- The 'gambler's fallacy' where people believe that if a particular outcome (e.g., a coin landing heads up) has occurred more frequently than expected in the recent past, it is less likely to occur in the future. Conversely, if an outcome has occurred less frequently than expected, it is believed to be more likely in the future. In reality, past events do not influence future outcomes when the events are independent and random.
- When creating phishing emails, scammers use fake URLs, text formatting, logos, and graphics that resemble legitimate ones, and the victims find it difficult to spot the differences and accept the emails as legitimate.

4.2.3 Confirmation Bias

Confirmation bias refers to the tendency of individuals to favor, interpret, or recall information in a way that confirms their preexisting beliefs or values. Conversely, it is challenging to persuade them that an idea is incorrect when it aligns with their established beliefs or values. Once the belief is established, individuals are more inclined to overlook inconsistencies in the data presented to them.

- Flat Earth believers may interpret visual observations or experiences in a way that confirms their belief in a flat Earth. For instance, they might selectively focus on horizon appearances that seem flat while dismissing scientific evidence such as satellite images.
- In a phishing email masquerading as a communication from PayPal, once the recipient begins to trust that the provided link will lead them to the legitimate PayPal site, their confidence extends to other elements. This false sense of security is reinforced by the clustering illusion, wherein the user unconsciously overlooks minor discrepancies in the URL, logos, graphics, and other details.

4.2.4 Zero-risk Bias

Zero risk bias refers to the tendency of individuals to prefer options or courses of action that minimize or eliminate a specific risk entirely, even when the overall reduction in risk is negligible or the alternative options might have more substantial benefits. In other words, people may be drawn to options that eliminate a perceived risk entirely, even if alternative choices with more nuanced risk-reward trade-offs might be more beneficial or practical.
E.g.:

- When asked 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.
- Developers tend to focus on absolute, information-theoretically secure, security solutions even if they are nearly impossible to deploy while ignoring solutions that achieve very good but not perfect security, even if they reduce overall risk.
- Developers implementing extremely complex encryption protocols for all communication channels, even for low-risk internal communications. While strong encryption is essential, an excessively complex system might introduce usability issues without proportional security benefits.

4.2.5 Biases and the Importance of Default and Nudges

Given the inherent biases in the human mind that introduce vulnerabilities into systems, leveraging defaults and nudges emerges as a strategic approach to enhance the security of these systems.

In systems, "**defaults**" refers to pre-established settings or configurations that are automatically applied if a user or administrator does not make explicit choices or selections. Individuals frequently opt for the path of least resistance, choosing the default or standard configuration of a system under the assumption that it should suffice. This tendency is often reinforced by present bias. Hence, security engineers can leverage defaults to configure users' systems to have expected levels of security.
E.g.:

- Have automatic updates for the system and installed software enabled by default.
- Have encryption for sensitive data or communication channels enabled by default.
- Configuring privacy settings to restrict data collection and sharing by default.
- Assigning minimal privileges to resources by default, with the option to grant additional permissions as needed.

At the same time, some service providers can misuse defaults. For instance, in most social networks, default settings are relatively open for information sharing unless a user goes into the settings and changes them. When many users enhance their privacy settings, the platform may alter its architecture/settings periodically, requiring users to revisit and opt out again. This takes advantage of hazardous defaults and underscores the **control paradox**, which refers to less privacy-conscious behaviour in users when provided with more control over their privacy settings.

"**Nudges**" refers to subtle, positive reinforcements or gentle prompts designed to shape user behavior by presenting information or options in a way that encourages a desired response. These interventions are typically non-coercive and maintain the users' ability to make their own choices. Security engineers can harness the power of nudges to prompt users to take actions that enhance system security seamlessly, without causing disruptions to their primary tasks. E.g.:

- Friendly notifications encouraging users to install the latest software updates.
- Gentle reminders to enable two-factor authentication.
- Prompts suggesting users review and adjust privacy settings.

While nudges can be highly beneficial, it is important to acknowledge that excessive use of nudges may lead to adverse outcomes stemming from **user conditioning**, as discussed in Section 6.

5 Techniques to Influence People

Biases in the human psyche can be methodically exploited to influence decision-making. Psychology professor Robert Cialdini describes six primary categories of techniques that can be employed to influence individuals in their decision-making process. These are closely related to the biases we described earlier, and for the scope of the class, you can also consider them as biases.

1. **Reciprocity**: The human tendency to reciprocate/return kindness or favors
   E.g.:
   - If a friendly caller, claiming to be from the IT department, guides a new employee through the usage of a new software and subsequently suggests installing a "new tool that IT is rolling out," there is a significant likelihood that the employee may proceed with the installation (of potential malware), even if it goes against company policies that prohibit such actions.
2. **Commitment and consistency**: The human tendency to maintain and stick to previously made decisions, beliefs, or commitments.
   E.g.:
   - When an attacker initiates a call with a new employee, purportedly providing information about security policies, the attacker induces the employee to make an explicit commitment. If, following this commitment, the attacker then requests the employee's password under the guise of ensuring compliance, the employee is more likely to disclose their password, driven by a sense of obligation to uphold their commitment. The deception can continue when, having obtained the initial password, the attacker further manipulates the user into selecting a supposedly 'better' password — one that the attacker can easily guess.
3. **Social proof/validation**: The human tendency to seek the approval of others and rely on the actions and opinions of others to make decisions.
   E.g.:

- If a caller contacts an employee, claiming to conduct a company-wide survey and mentions names of other employees in the same department who have supposedly taken part in the survey, the likelihood of the victim participating in the survey and revealing sensitive information significantly increases.
- In the 2020 Twitter Bitcoin scam, attackers hijacked and used verified Twitter accounts of several high-profile personnel like Elon Musk, Bill Gates, and Barack Obama which encouraged people to take part in the fake donation.

4. **Like bias**: The human tendency to comply with requests coming from people they like due to their natural desire to be liked by them as well.
   E.g.:
   - An attacker strategically befriends a victim, professing a shared enthusiasm for a popular pop singer, whom the victim is an avid fan of. After establishing this connection over a few days, the attacker requests the victim to install and try out a new software they claim to be developing. In this scenario, the victim is highly likely to accept the request, influenced by the perceived shared interest and trust in the newfound friendship.

5. **Respect to authority**: The human tendency to defer to or obey figures of authority, often without critical evaluation or independent judgment.
   E.g.:
   - In a 2006 experiment conducted by R. Cialdini in a hospital setting, nurses received a call from a person claiming to be a physician in the hospital requesting them to administer medication to a patient. In 95% of the instances, the nurses complied with the request, influenced by the perceived authority of the caller.

6. **Scarcity**: The human tendency to place a higher value on items that are perceived as being in short supply from a fear of missing out.
   E.g.:
   - When a scammer utilizes SMS to advertise a highly desirable, '*limited-edition*' product falsely claiming that only a few items are left and prompts the user to click on a provided link for purchase, the probability of the user succumbing to the scam is significantly high.

6 User Conditioning

In addition to the techniques mentioned in Section 5, user conditioning is another key factor that can significantly impact individuals' decision-making processes.

User Conditioning refers to the process of getting users habituated to react to certain situations in a specific way ('click-whirr reaction') through repeated exposure, reinforcement, or manipulation of stimuli. Consequently, when confronted with a security decision, users often react automatically instead of engaging in thoughtful and deliberate decision-making.
E.g.:

- When users are consistently bombarded with warning messages, they may develop a conditioned response to hastily dismiss and *click away* these alerts without reading the warning or thoroughly considering the potential risks or implications.

- In the context of the Windows Recycle Bin and the option to delete files permanently, when users repeatedly encounter the prompt asking if they are sure they want to delete a file permanently, users may develop a habitual response of confirming permanent deletion without thinking. This habitual response increases the risk of unintentionally deleting crucial files, rendering the warning ineffective despite its presence.
- Many certificate errors that trigger warnings are typically benign and haven't been associated with any practical attacks. Consequently, users often habitually dismiss these warnings and proceed to visit websites without experiencing any adverse consequences. This conditioning can lead users to ignore warning messages even for potentially harmful certificate errors, as they become accustomed to the belief that such alerts are generally inconsequential.

7 User Education

Efforts to utilize user education in mitigating vulnerabilities within security solutions stemming from human behavior have yielded mixed results. The mere dissemination of company policies has proven ineffective in achieving positive outcomes. However, training approaches that incorporate feedback and iterative processes have demonstrated more favorable results.

While user education can contribute to mitigating vulnerabilities resulting from human behavior, it is crucial to acknowledge its insufficiency. A significant limitation is the lack of user comprehension of the 'security language' employed by security engineers and developers. Additionally, users often lack intrinsic motivation to delve into security intricacies, given their focus on primary job responsibilities. Hence, as security engineers, the imperative is to design inherently intuitive systems, seamlessly integrating the ideal security solutions to the extent that users are unaware of their presence.

8 Deception in Practice

Next, we explore two real-world examples that exploit vulnerabilities in human psychology within the context of security engineering.

8.1 Social Engineering

Social engineering targets the most vulnerable link in the cybersecurity chain — humans, to disclose confidential information, grant unauthorized access, or perform actions that compromise the security of a system to which they belong. Social engineering is especially useful in stealing credentials and identities.
E.g.:

- An old scam from the 20th century involved stealing someone's ATM card and subsequently posing as a bank representative during a phone call to extract the PIN. The attacker would inquire if the victim's card had been stolen, and upon receiving an affirmative response, the attacker would imply the necessity of the PIN to cancel the card. This manipulation often led the victim to unwittingly disclose their PIN.
- In 2007, the Treasury Inspector General for Tax Administration conducted an audit of the Internal Revenue Service (IRS). During this audit, 102 IRS employees at various

levels were contacted by the audit staff, who requested their user IDs and instructed them to change their passwords to a pre-determined value. Surprisingly, 62 of the contacted employees complied with this directive.

8.2 Phishing

Typically executed via deceptive emails, messages, or websites that appear to be from reputable sources, phishing aims to trick recipients into providing confidential information, which can then be exploited for fraudulent activities, unauthorized access, or identity theft. While phone-based social engineering held prominence in the 20th century, online phishing has emerged as the predominant tactic to hack systems through people in the 21st century.
E.g.:

- 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 authorities.
- Recent phishing attacks against banks often reuse genuine bank emails, with just the URLs changed for fake login windows.

9 Usability and Security of Passwords

The management of passwords provides an insightful intersection of usability, applied psychology, and security. Considering the limitations of human memory—difficulty remembering infrequently-used or frequently-changed items, the inability to forget on demand, the challenges of recall compared to recognition, and the increased difficulty with non-meaningful words—it becomes evident that passwords might not be the optimal choice for authentication.

9.1 Main Concerns related to Passwords

There are three broad concerns related to passwords:

1. Will the user enter the passwords correctly with a high enough probability?
2. 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?
3. 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?

9.1.1 Password entry

If a password is too long or complex, users might have difficulty entering it correctly. Since 2010, common password rules have mandated the inclusion of at least one lower-case letter, upper-case letter, number, and special character. Entering such a password could be cumbersome and intricate, particularly when accessing systems via mobile phones.
E.g.:

- The firing codes for US nuclear weapons consist of only 12 decimal digits as experiments suggested that 12 digits was the maximum that could be conveyed reliably in challenging

circumstances where operators will be under extreme stress, and possibly using improvised or obsolete communications channels.

Users typically prefer using simple passwords, to facilitate easy entry and minimize errors.

9.1.2 Remembering the password

A significant drawback associated with passwords is the challenge users face in remembering them. When individuals are tasked with memorizing passwords, they often opt for easily guessable values or resort to writing them down, sometimes adopting both practices. Traditional password advice encapsulates the paradox: "Choose a password you can't remember, and don't write it down." However, the appropriateness of writing down passwords varies with the context. In an open office environment or public space, writing down a password is strongly discouraged. Conversely, in a private space like one's own bedroom, it may be advisable to select a complex password and keep it securely written down.

9.1.3 Disclosing the Passwords

An attacker can extract passwords from users through various malicious attacks such as social engineering and phishing or they can get it during password entry or storage.

Password entry: Attackers can exploit vulnerabilities in interface design to acquire users' passwords during password entry. For instance, in public EFTPOS scenarios, some customers fail to cover their PINs when entering them, providing an opportunity for attackers to observe and capture the PIN. Additionally, attackers can also exploit technical defeats of password retry counter through timing attacks.

Password storage: If attackers manage to access stored password lists (typically hashed and salted), they can leverage techniques such as brute-force attacks, dictionary attacks, or various password cracking tools to reveal the passwords. Additionally, people use password managers as tools to securely store their passwords. However, these tools have limitations, as outlined in Section 9.3.

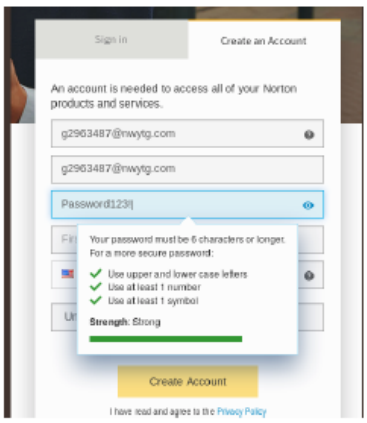
9.2 Outdated Advice

- Passwords must be changed frequently (e.g.: every 30/60/90/180 days)
  In general, this is bad advice for two reasons.
  - If stored correctly, strong passwords remain strong and hence do not need to be changed.
  - Humans can remember a handful of strong passwords at best. Therefore, when forced, they will choose ever-weaker passwords, and passwords for different logins become more similar over time when they run out of ideas for memorable passwords.

  Consequently, better advice would be '**Do not expire passwords unless you suspect they have been compromised**'. UK policies have recommended this since 2015, and Australia followed in 2017.

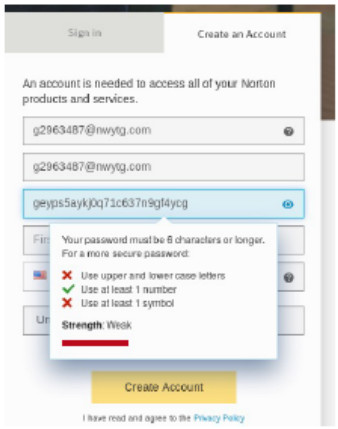
- Complexity rules
  Blindly following a complexity rule like '**Your password must consist of 6-12 characters, with at least one special character and two numbers**' will not always result

in strong passwords. As shown in Figure 2, this rule will accept a weak password like '`Password123!`' while rejecting a strong password like '`geyps5aykj0q71c637n9gf4ycg`'.



Source: https://github.com/OCA/server-auth/issues/111

(a) Weak password accepted



Source: https://github.com/OCA/server-auth/issues/111

(b) Strong password rejected

Figure 2: Complexity Rules

An ideal password should still be,

- **Unpredictable** which implies complexity and length
  However we already know that checking for complexity with a complexity rule alone is bad advice. A much better approach is to,
  - 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.

  The reasoning behind the aforementioned advice is that brute-forcing is an expensive and resource-intensive method for attackers, typically employed as a last resort. Instead, attackers often opt for dictionary attacks first, as they are more cost-efficient.
- **Memorable** which causes tension to memorability
  To tackle the memorability challenge associated with lengthy and complex passwords, the following methods have been tried and tested.
  - Password managers
  - Passphrases via Dice method (or similar)
  - 'First letters from a sentence' method (but careful)
  - PAO method (but careful)

9.3 Password Managers

A password manager is a software application or service designed to securely store and manage users' passwords and other sensitive login information. It provides a convenient and secure way for users to store their various passwords for different accounts, such as email, social media, banking, and more. Some popular password managers include *Apple Keychain*, *Google Chrome*

Password Manager, *Firefox Lockwise*, *Bitwarden* and *LastPass*. The main functionalities of a password manager include:

- The database of passwords and logins is protected by a strong master password.
- Can generate highly complex and long passwords
- Often integrated with browsers and other applications
- Can copy and paste passwords into the application
- Some password managers are available for mobile OSes
- Passwords are synchronized across multiple devices

While password managers streamline the utilization of lengthy and intricate passwords alleviating concerns about memorization, they also present inherent drawbacks.

- Loss of master password means loss of all passwords
- Portability remains an issue as it is not widespread on mobile devices
- Longer workflow
- Some sites prevent copy and paste (bad idea !!)

Several studies have examined the risks and usability of password managers. However, there appears to be a lack of recent research specifically focused on evaluating the usability of modern password managers.

10 Creating Secure and Memorable Passwords

As previously discussed, a strong password should be complex, long, and memorable. Contemporary recommendations advocate for the use of passphrases comprising three or more randomly selected dictionary words to fulfill these criteria. Empirical research indicates that users tend to opt for multi-word passphrases with significantly lower entropy compared to genuinely random selections from a dictionary. Typically, users favor common noun bigrams, and expanding to three or four words results in diminishing returns concerning security benefits. The following techniques are proposed to select random words and construct a secure passphrase.

10.1 Passphrases via Dice method

The main concept behind the dice method is to randomly choose 7-12 words from the dictionary to create the passphrase. Leveraging words enhances memorability, particularly when linked with a story. Furthermore, the passphrase can be further improved by methods such as replacing words with characters, extending its complexity and security. The following steps outline how the process works.

1. Get a word list: Obtain a word list with a list of unique words associated with each possible combination of five dice rolls. This list typically contains thousands of words, each assigned to a unique five-digit number. (E.g.: EFF's Long Wordlist)
2. Roll the dice: Roll five standard six-sided dice. Each dice roll generates a random number between 1 and 6. Record the result as a five-digit number.

3. Match the number to the word list: Match the result of the dice rolls to the corresponding numbers on the word list. Each set of five dice rolls corresponds to one word on the list.
4. Repeat for the desired number of words: Repeat steps 2 and 3 as many times as needed to create a passphrase with the desired number of words.
5. Combine the words: Once you have chosen the desired number of words, combine them to form your passphrase.

While passphrases generated by the Dice method offer reasonably high security, they do entail a significant amount of typing, which can be seen as a limitation.

Dice Method Example

Suppose you generated the following sequence of numbers by throwing a dice 30 times.

54452 61644 63545 53161 42142

By referring to the EFF's Long Wordlist we can obtain the word sequence:

sizable tapioca undoing scalded nemesis

Now, either we can remember this word sequence or create a mnemonic to remember the phrase. It might be a story, scenario, or sentence that we will be able to remember and that can remind us of the particular words we chose in order. In general, people are better at remembering sentences than at a random sequence of numbers.

For example, "The **sizable** bowl of **tapioca** pudding was **undoing** my patience as I **scalded** my tongue, cursing my **nemesis**."

Note: The word list for the Dice method must have $6^5 = 7,776$ words, which is the number of possible numbers that can be obtained by rolling a dice 5 times.

10.2 Letters-from-a-Sentence

For this method, a user chooses a memorable sentence and picks one to two letters from each word to create a password.
E.g.: Wow! 62 students, all in this memorable class of 2024! –> W!62s,aitmco2025!

- Ideally, has upper and lower case words.
- Ideally, has digits (e.g. dates, amounts).
- Ideally, has special characters
- Can be easily extended: W!62s,aitmc02o25!

10.3 Person-Action-Object (PAO) Method

The premise of the PAO method is to leverage visual cues and unusual imagery to stimulate memory recall. The user begins by selecting a random place and person and then visualizes a random action and object involving these chosen elements. Subsequently, the user may

establish rules to transform this constructed scenario into a password.

E.g.: If the user chose *Dath Vader* and *Mt Everest* as the person and place, and *riding* and *pony* as the action and object, the resulting sentence could be,

Darth Vader riding a pony on Mt Everest.

One possible password could be *DVrapoMtE!* which can be easily extended to *DVr4poMt3!*

11 Password Complexity Calculations

11.1 Combinatorics: Calculating the Number of Possible Passwords

Password complexity can be measured by calculating the total number of possible passwords using combinatorics. This involves determining the number of different characters that can be used in each position of the password and then calculating the total number of combinations.

Example 1: Simple Password Calculation

- **Length of Password (L)**: 6 characters
- **Character Set (C)**: Lowercase letters (26 characters)

The number of possible passwords N can be calculated as:

$$N = C^L = 26^6$$

This gives:

$$N = 26^6 = 308,915,776$$

So, there are 308,915,776 possible 6-character passwords using only lowercase letters.

Example 2: Complex Password Calculation

- **Length of Password (L)**: 8 characters
- **Character Set (C)**: Uppercase letters, lowercase letters, digits, and special characters ($26 + 26 + 10 + 10 = 72$ characters)

The number of possible passwords N can be calculated as:

$$N = C^L = 72^8$$

This gives:

$$N = 72^8 \approx 7.2 \times 10^{14}$$

So, there are approximately 720 trillion possible 8-character passwords using a diverse character set.

11.2 Entropy: Measuring Password Strength

Entropy is a measure of the uncertainty or randomness in a password, usually expressed in bits. Higher entropy indicates a more complex and secure password.

Entropy Calculation Formula:

$$H = \log_2(N)$$

Where:

- H is the entropy in bits.
- N is the total number of possible passwords.

Example 1: Entropy of a Simple Password Using the earlier example with 6 lowercase letters:

$$N = 26^6$$

$$H = \log_2(26^6) = 6 \log_2(26)$$

$$H \approx 6 \times 4.7 = 28.2 \text{ bits}$$

Example 2: Entropy of a Complex Password Using the example with 8 characters from a set of 72:

$$N = 72^8$$

$$H = \log_2(72^8) = 8 \log_2(72)$$

$$H \approx 8 \times 6.17 = 49.36 \text{ bits}$$

11.3 Password Complexity Examples

Example 1: Simple Password

- Password: `abcdef`
- Length: 6
- Character Set: 26 lowercase letters

Number of possible combinations:

$$N = 26^6 = 308,915,776$$

Entropy:

$$H = \log_2(26^6) \approx 28.2 \text{ bits}$$

Example 2: Moderate Complexity Password

- Password: `a1B!9xYz`
- Length: 8
- Character Set: 26 lowercase + 26 uppercase + 10 digits + 10 special = 72 characters

Number of possible combinations:

$$N = 72^8 \approx 7.2 \times 10^{14}$$

Entropy:

$$H = \log_2(72^8) \approx 49.36 \text{ bits}$$

Example 3: High Complexity Password

- Password: `a1B!9xYzA1B!`
- Length: 12
- Character Set: 26 lowercase + 26 uppercase + 10 digits + 10 special = 72 characters

Number of possible combinations:

$$N = 72^{12} \approx 3.7 \times 10^{22}$$

Entropy:

$$H = \log_2(72^{12}) \approx 74.1 \text{ bits}$$

Summary

- **Combinatorics**: Calculate the total number of possible passwords by raising the size of the character set to the power of the password length.
- **Entropy**: Measure the strength of a password in bits by calculating the logarithm base 2 of the total number of possible passwords.

Dice Method - Passowrd Complextiy Calculation

These days, it is recommended to use six dice words for password generation.

Let's assume the following for the complexity calculation.

- The attacker knows the wordlist used
- The attacker knows how many words were used to build the password.

As we discussed earlier, there are $6^5 = 7,776$ word possibilities. Therefore, the total six-word possibilities are $7,776^6 \approx 2.2107392 \times 10^{23}$.

The entropy is $\log_2 2.2107392 \times 10^{23} \approx 77.55$ bits.

12 Password Practicalities

Let's summarizes the practicalities associated with using strong passwords.

- Different passwords for every site?
  - **In theory, we should use different passwords for every site** - The methods employed by websites to safeguard passwords are not always transparent, and a breach on one site can potentially expose both the username (often an email address) and password. If these credentials are reused on other sites, attackers could gain unauthorized access to those accounts as well.
  - In practice, most of us have dozens of accounts and it is not possible to memorize them all.

  - Password managers offer an excellent solution to this issue, particularly if you don't want to carry all your passwords with you at all times (for mobility reasons) and if you possess a solid understanding of how password managers function.

- 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 is 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
  - Writing down is **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.

- Passwords will be Outdated?
  - Alternatives to passwords
    * **SMS**: Mostly used as a secondary authentication factor. It is a legacy technology and, as a result, can be spoofed or intercepted easily.
    * **Physical authentication tokens:** Chip authentication program (CAP) reader (enables logging on to online banking with bank cards), and authentication tokens used by companies like Google and Microsoft to authenticate staff are examples.
    * **Biometrics:** Biometrics provide a reliable and easy-to-use alternative to passwords. The effort required to use biometrics for authentication is minimal. But often, biometrics alone have limitations. They can be spoofed, can change over time or health conditions, and if compromised, they are gone for good. The most viable approach is to use biometrics as another factor in the authentication process. For example, most of the time, you unlock your phone by using the faceID, but periodically, your phone also prompts you to enter the PIN. The same goes for laptops, where users are asked to to enter the password in some scenarios (e.g., after a reboot).

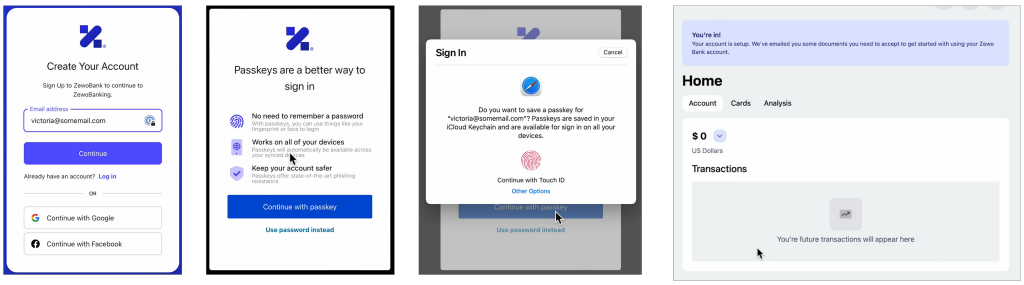
Passwords will be there for a few more years and will remain a usability challenge for the user and a security challenge for security engineers. However, a few passwordless authentication technologies are emerging, and some online services and apps are adopting them. Naturally, these solutions are taking the advances in usable security and, as a result, are more usable while being secure.

13 Passkeys

Many web services are now integrating "passkeys" as a form of authentication to their systems. When accessing a website, users authenticate their identity through biometric verification on their devices rather than entering a password.

13.1 User Registration

Overall, the user registration process using passkeys is shown in Figure 5. The user doesn't have to do anything besides give the email address and get locally authenticated using biometrics.



1) User enters the email address

2) User selects the passkey option

3) User is locally authenticated using biometrics

4) User is logged in

Figure 3: Biometric passkey authentication process [?]

While the user doesn't seem to do anything much, in the background, a few things are happening.

Passkeys - Registration Process

a) User Initiates Registration: The user starts the registration process on the relying party's website or service.

b) Challenge Issued: The relying party (i.e., the website at the other end) generates a challenge, a random piece of data, and sends it to the user's device.

c) Biometric Verification: The user verifies their identity using a biometric method (e.g., fingerprint or facial recognition) or another form of local authentication (e.g., PIN).

d) Key Pair Generation: Upon successful verification, the authenticator (i.e., the device platform) generates a unique public-private key pair. Private Key Storage: The private key is securely stored on the user's device (in a secure enclave or TPM).

e) Public Key Registration: The public key, along with the challenge response (signed challenge using the private key), is sent to the relying party. The relying party stores the public key for future authentication.

13.2 User Verification

The login process is as simple as the user authorising the use of the stored passkey using the device's biometrics, as shown in Figure 4.

Passkeys - Registration Process

a) User Initiates Login: The user tries to log in to the relying party's website or service.

b) Challenge Issued: The relying party generates a new challenge and sends it to the user's device. Biometric Verification: The user again verifies their identity using a biometric method or another form of local authentication.

c) Challenge Signing: The authenticator uses the stored private key to sign the challenge. Response Transmission: The signed challenge (response) is sent back to the relying party.

d) Verification: The relying party uses the stored public key to verify the signed challenge. If the verification is successful, the user is authenticated.

This process is schematically shown in Figure 5.

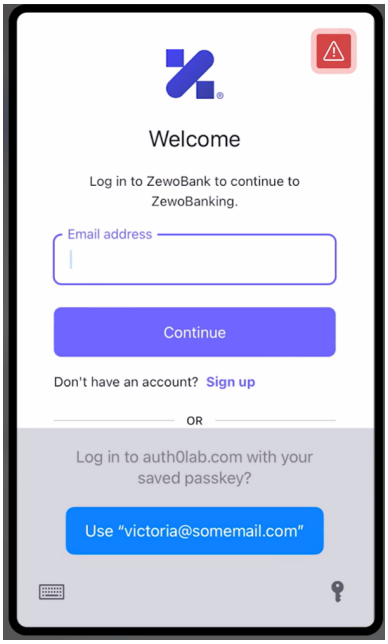


Figure 4: Biometric passkey verification process [?]

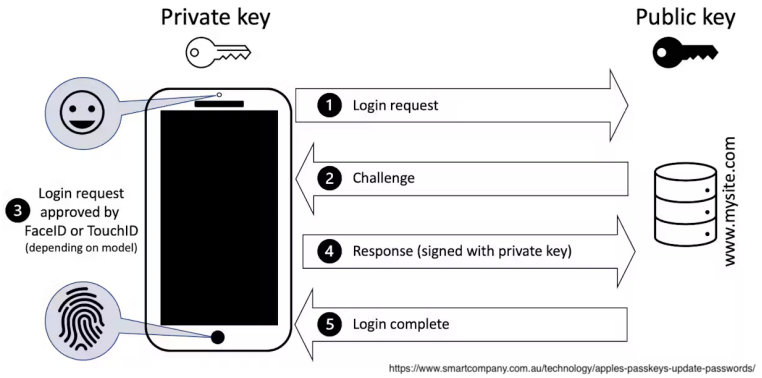


Figure 5: Biometric passkey authentication process [?]

14 UX/UI Designs for Security

14.1 Mental Models

"Mental models" refer to the internal representations or frameworks that individuals develop to understand and interact with systems, objects, or processes. These models are based on their experiences, beliefs, and interpretations, and they shape how users perceive and navigate through the system. In the context of user experience design and human-computer interaction, understanding users' mental models is crucial for creating intuitive and user-friendly interfaces. Several companies like Apple and Google have become very proficient in exploiting mental models. Similarly, designers of security solutions can leverage these mental models to guide users toward making secure decisions within the system.

14.2 Concepts from UX/UI Design

14.2.1 Affordance

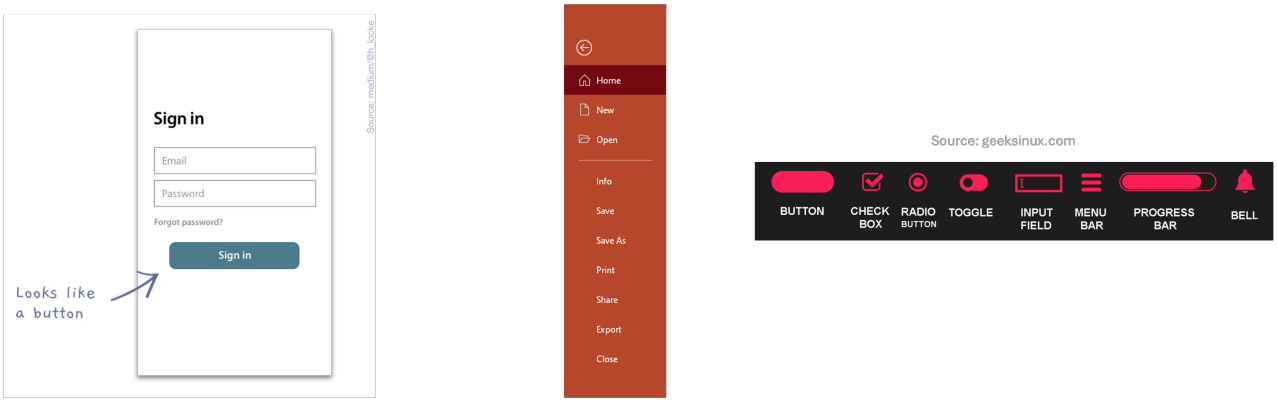
Affordance refers to design aspects that guide the user towards the correct behaviour. In security engineering, it could include visual cues or interactive elements that suggest how users can interact with security features or guide them to use security features. For instance, in Figure 6(a), when perceiving the grey *Sign in* button on the screen, users perceive that they need to take further action before it can be pressed. Then, once the form is completed, the button turns green, and the user perceives that it can be pressed to produce an action.

14.2.2 Constraints

Constraints are limitations or restrictions that guide users' interactions with a system or object. In the context of security solutions, constraints may include password length requirements, restrictions on the types of characters allowed in passwords, or providing only a limited number of options. These constraints help enforce security measures and protect against unauthorized access. For instance, the panel in Figure 6(b) still allows users to move around and interact but are confined to a certain area under certain parameters.

14.2.3 Conventions

Conventions refer to established norms or standards that users expect to encounter based on their prior experiences with similar systems or interfaces. For instance, in Figure 6(c), the symbols used in the panel are standard symbols users are familiar with.



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

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

c) Conventions: Users can recognise the elements and their functionalities

Figure 6: Examples of UX design concepts process [**?**]

Mental Models and UX Concepts

Affordances leverage users' mental models to indicate possible actions and guide correct usage.

Constraints shape and refine mental models by limiting possible interactions.

Conventions align with mental models based on familiar patterns and norms, making interfaces predictable and easier to use.

15 Takeaway Messages

In conclusion, the effectiveness of security features relies heavily on thoughtful and user-centric design. Blaming users for security failures is unproductive; instead, designers must acknowledge that users do not share the same level of technical expertise or stress tolerance as developers. It is crucial to incorporate psychological considerations into the design of new security features to ensure they are intuitive and stress-free. Adopting a design process that involves stakeholders' perspectives can lead to more robust and user-friendly security solutions. Relying solely on user education is insufficient and should not replace the fundamental principles of good design and user engagement. By following these guidelines, security measures can be more effective and better integrated into users' daily interactions.

16 Practice Questions

Question 1: A while ago, one of your friends received a phone call from an unknown number. The caller said they were from the Australian Taxation Office (ATO) and had an issue with your friend's tax return. First, the caller wanted to verify your friend's identity and asked questions like your friend's legal name, address, TFN, and job information. A few months later, your friend was found to be a victim of identity theft.

What bias in the human psyche was the attacker exploiting to get information from your friend?

a) Confirmation bias

b) Reciprocation bias

c) Respect to authority

d) Like bias

Explanation: *The correct answer is c). The caller claims from ATO and asks some official-sounding questions. Therefore your friend is likely to comply under the bias of "respect to authority".*

"Confirmation bias" is the tendency to interpret new evidence as confirmation of one's existing beliefs or theories. While attackers can exploit this, the given scenario doesn't directly use confirmation bias. But there is a slight relationship here. The second step of asking for information to verify can be considered a case of using confirmation bias. If the target is already convinced that the call was from ATO, the second step adds to the confirmation bias. However, here "respect to authority" is a more direct answer.

"Reciprocation bias" bias describes the impulse to reciprocate actions others have done towards us. Scammers can use this (for example, by providing a small gift or by making it appear they are bending the rules in the recipient's favour) to successfully pitch their scams.

"Like bias" means that we overvalue people who are like us, or friendlier towards us or who we think are like us. Similar to reciprocation, this tendency can be exploited by attackers by especially appearing as friendly and approachable.

Question 2: You have implemented the following password rules:

- A password must be 5 or 6 characters long
- A password can contain
  - . [a-z] lowercase letters
  - . [0-9] numerical characters
  - . Special characters * or @

What is the total number of possible passwords if no character repetition is allowed?

a) 1,203,322,288

b) 7,624,512

c) 2,760,681

d) None of the above

Explanation: *The way to calculate this is by using combinations. We can separately calculate the number of all possible five-character passwords and the number of all possible six-character passwords.*

There was a mistake in this question. It should have mentioned repetition of characters are not allowed.
The number of choices for a character = 26 (lower case characters) + 10 (digits) + 2 (or @) = 38 The number of 5-character passwords = 38P5 = 60,233,040*

The number of 6-character passwords = 38P6 = 1,987,690,320

Total number of possible passwords = 2,047,923,360

Therefore, the answer is none of the above.

Question 3: Which of the following are true about the usability and psychology principles we learned in the class? Select all that apply.

a) In most of the current environments, it is rational for users to ignore security warnings.

b) Affordance is the idea of designing systems where users are constrained from making errors.

c) The principle of psychological acceptability says it is essential that the human interface is designed for ease of use so that users routinely and automatically apply the protection mechanisms correctly.

d) The principle of psychological acceptability says that a security mechanism should not make a resource more difficult to access than if the mechanism were not present.

Explanation: *In class, we had two definitions of the principle of psychological acceptability, which are given in c) and d). The original definition from the Saltzer and Schroeder paper is given in c). A simpler definition we discussed is given in d). a) Is correct because users often ignore warnings to get their work done or because of user conditioning.*

b) is incorrect because it is not the definition of affordance. Affordance is the idea of guiding the user towards the correct behaviour. Therefore the correct answers are a), c), and d).

Question 4: What techniques do attackers use to find a login password of a target user? Select all that apply.

a) Checking social media profiles and trying to guess passwords

b) Dictionary attacks

c) Brute-force attacks

d) Social engineering

Explanation: *All the answers are correct.*

a) Especially when conducting targeted attacks, the attacks comb through the target users' social media profiles to find information such as birthdays, pets' names etc.

b) Dictionary attacks try the most commonly used passwords one after the other, assuming the target user is using one of the most commonly used passwords. Most commonly used passwords usually include passwords such as "password", "123456", "123456789", "guest", and "qwerty".

c) Brute force attacks try all possible combinations of characters to obtain a match. This usually works well for short passwords or if the attacker knows parts of the passwords.

d) Social engineering attacks involve the attacker interacting with the target user either via email, in person, or over the phone and trying to trick the user into providing the password or parts of it.

Question 5: We are going to generate a password using the Diceware method.

Read about it here `https://www.eff.org/dice` Suppose you are generating a password using the EFF large word list

`https://www.eff.org/files/2016/07/18/eff_large_wordlist.txt`

and you rolled the following numbers.

32345 21345 25662 12456 11234 23456

What is your password going to be?

a) gloomily creative everglade attempt acts dispatch

b) active bonus extent relocate dispense fax

c) gloomily fantasize idiom polar simmering specimen

d) None of these

Explanation: *Answer is a). Diceware method allows the creation of random yet memorable passwords. In general, humans are good at remembering words than numbers. A sequence of words can be memorised by relating it to some 'story'. In this case, you have given the randomly generated numbers. So you have to look up the word list to get the corresponding words.*

Question 6: I am trying to generate a password using the PAO method. I have thought about the following scenario

Harry Potter is eating Sushi at Broadway

What is a password I can generate using this?

a) HaPoiseaSuatBr

b) ry3r1sngh14t4y

c) HPieSaB

d) All of these

In PAO (Person-Action-Object) method, the idea is to imagine a random person and a random action on a random object involving the person. Once you form the sentence, you can have your own rule to build the password, such as "taking the first two letters of each word", "last two letters of each word", etc.

a) is correct because it takes the first two letters of each word.

b) is correct because it takes the last two letters of each word but with some additional rules. 'E'e is replaced by '3', 'i' is replaced by '1', and 'a' is replaced by '4'.

c) is correct because it takes the first letter of each word.

Therefore the correct answer is d) All of these.

Question 7: You work in a government institution that handles classified information. You take public transport to work. There is this person who sometimes takes the same train as you, and you two start talking. This person buys you coffee from time to time, and you two get along well and like each other. You notice that this person is subtly asking questions about your team at work, information about your bosses, and the general layout of the office building.

What bias/biases in the human psyche is this person trying to exploit to obtain information about the government facility? Select all that apply.

a) Social validation

b) Confirmation bias

c) Reciprocation bias

d) Like bias

"Social validation" is humans follow or conform to the actions of others within a group. We

are more likely to do something if we are told that some of our office colleagues have already done that. This is not applicable in this case.

"Confirmation bias" is the human tendency to search for, interpret, favour, and recall information in a way that confirms or supports one's prior beliefs or values. Again not applicable in this case.

"Reciprocation bias" is. Here the person buys you coffee, and therefore you feel obliged to do something in return. Therefore it is correct.

"Like bias" is the human tendency to respond positively to the requests of people they like or are more amicable. In this case, the text says the two people like each other. Therefore "Like bias" is correct.

Question 8: Following is a list of domains that are registered on the web. As you can see, they look very close to the popular domain name "google.com". What kind of human error is the attacker trying to exploit here?

Figure 7: Operation of One-Time Pad

a) Failure of a manual skill

b) Following the wrong rule

c) Cognitive reasons

d) All of them

Explanation: *The question is related to the three categories of human errors discussed in class. Failure of manual skill, in this case, refers to the fact that some users might make mistakes while they are typing google.com in their browser's address bar. Accidentally they might make a mistake in one character and enter a URL that is shown in red. So a) is the*

correct answer. However, let's check the other ones as well, just to be sure.

Following the wrong rule usually refers to errors in human understanding. For example, users might have the impression that the secure padlock sign next to a URL means the website is secure. However, setting up a malicious website with an SSL certificate is quite easy. So b) is not a correct answer.

Cognitive reasons refer to the fact that users may not understand the problem at all or biases of the human psyche force errors. Here a case can be made for something like a "clustering illusion" and say that the user sees a familiar pattern. But between a) and c), a) is a more related and direct answer.

Therefore the correct answer is a).

Question 9: Is writing down passwords always a bad idea? TRUE or FALSE

a) True

b) False

There are situations where writing down passwords don't pose much significant risk. For example, writing down your website password on your home study desk can be acceptable (given that the family members are trustworthy). Most of the time, attacks on web-mail or social media come as remote attacks. However, writing down in a more public place, say at your office desk, is not acceptable. Therefore the answer is "False", i.e., there are occasions where writing down the password is acceptable.

Question 10: It is a good practice to change passwords every three months. TRUE or FALSE

a) True

b) False

Explanation: *This is not always good practice. The main reason is that if the password is strong enough, there is no need to change it if there is no evidence that it has been compromised. Frequent password changes will be a burden to the user, and often the users will try to get around the system by making predictable changes. Therefore the answer is False.*

References

[1] Kate Conger and Nathaniel Popper. Hackers tell the story of the Twitter attack from the inside. *The New York Times*, July 2020. Accessed: 2026-08-07.