



# Human Factors in Cybersecurity

**Riccardo Scandariato**

Institute of Software Security, TUHH, Germany

ric\*\*\*do . scanda\*\*\*to @ tuhh.de

Master Course “Secure Software Engineering”

Summer Semester 2022 – Lecturer: Nicolás E. Díaz Ferreyra



# Agenda

- 1. Social Engineering**
- 2. Online Self-Disclosure**
- 3. Privacy Nudges**
- 4. Multiparty Privacy Conflicts**
- 5. Ethics**

# Social Engineering: Definitions

*“The ‘art’ of influencing people to divulge sensitive information”*

*“The science of using social interaction to persuade an individual to comply with a specific malicious request”* [Mouton, 2016]



The request, the persuasion, or the social interaction involve a computer-related entity.

# Social Engineering (SE)

SE refers to a BROAD range of malicious activities accomplished through **simple human interaction** and a fair amount of **deception**:

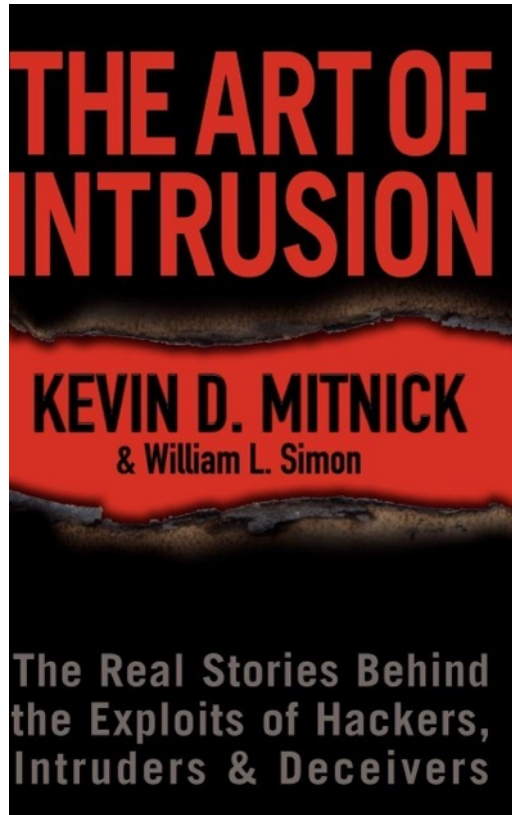
- People are the weakest security link of an organization.
  - *It's often easier for cybercriminals to manipulate a human than a computer network or system.*
  - *Attacks can be relatively low-tech, low-cost, and easy to execute.*
- Attackers use psychological manipulation to trick employees into making security mistakes or giving away sensitive information.
- **No one is immune!** Many smart and careful people can fall victim to a social engineering attack without even realizing it until it is too late.

Social Engineering can have **severe consequences** for businesses, financial institutions, and population as a whole.

# Getting Motorola's Source Code?









# Social Engineering

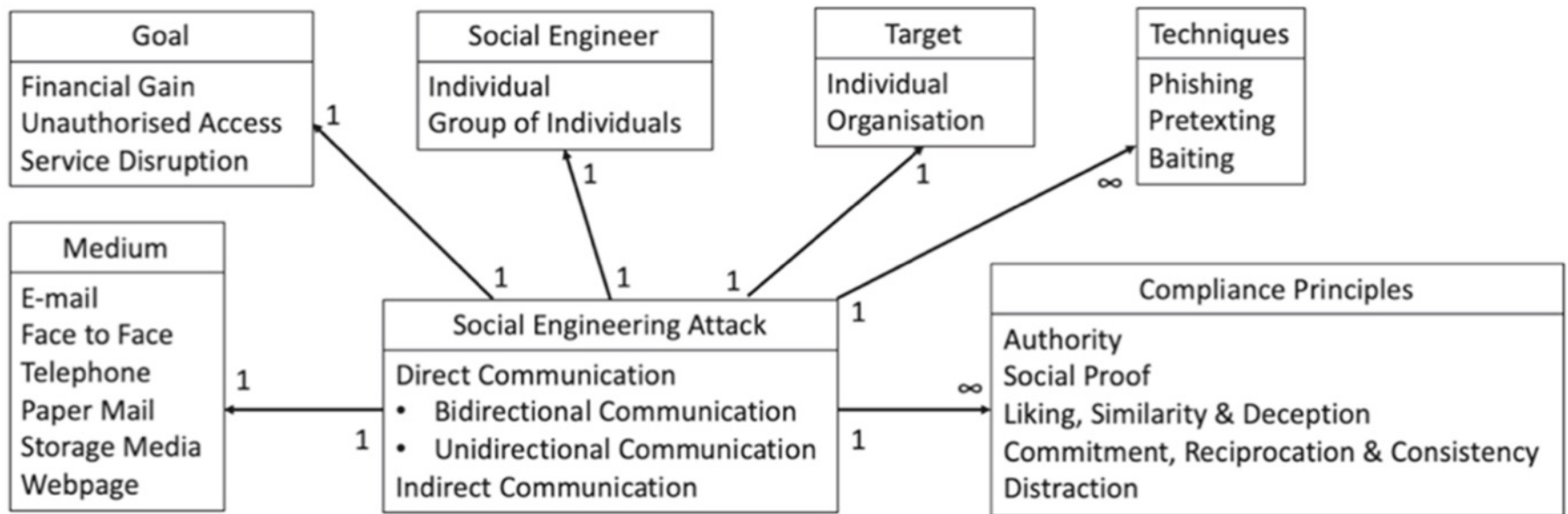


Kevin Mitnick is often considered the **original master of social engineering**. There are even books about (and authored by) him.

# Common Social Engineering Techniques

-  **Pretexting:** The attacker creates a scenario where the victim feels compelled to comply under false pretenses.
-  **Phishing:** The attacker sends *fraudulent emails*, claiming to be from a trustworthy source.
-  **Vishing and Smishing:** Same as phishing but using voice calls and text messages, respectively.
-  **Shoulder Surfing:** Use direct observation techniques to get information, such as looking over someone's shoulder at their screen or keyboard.
-  **Waterholing:** The attacker infects specific websites with malware and expect that some of their target companies' employees will visit them.
-  **Baiting:** Making false promises to users in order to lure them into revealing personal information or installing malware on the system.

# Social Engineering Attack Vector



At its core, a SE attack consists of a **Medium**, a **Goal**, a **Social Engineer**, a **Target**, plus one (or more) **Techniques** and **Compliance Principles**.



# Social Engineering Attack Cycle

The Social Engineering **attack cycle** comprises the following stages:

- 1. Attack formulation:** Identify the *goal of the attack* (e.g., financial gain) and the right *target* (e.g., individual).
- 2. Information gathering:** Collect information about the potential target and everything related to the attack.
  - *The sources can be anything or anyone with access to the information required for the attack.*
  - *Dumpster diving: Scan trash items for personal information.*
- 3. Preparation:** The social engineer analyzes the information and develop an action plan (i.e., an attack vector) to approach the target.

A target is more likely to share information with the attacker if a relationship exists between the two.

# Social Engineering Attack Cycle

4. **Develop a relationship:** The social engineer establishes a line of communication with the target and begin to build a relationship.
  - *If trust cannot be established, the required information is unlikely to be elicited from the target!! A good pretext simplifies this step 😊*
5. **Exploit the relationship:** The attacker employs **manipulation tactics** to get the target in a **desired emotional state** (e.g., as feeling sad or happy)
  - *The goal of emotional priming is making the target to feel comfortable about giving out information (and not guilty about it).*
  - *Once the target is in the desired emotional state, she can be exploited to obtain the necessary information (e.g., password).*
6. **Debrief:** The social engineer stays connected with the victim for a while, so she does not get alarmed/suspicious and contact the authorities.

# Desired Emotional States

Emotional State	Example
<b>Fear</b>	You receive a notification that you're under investigation for tax fraud and you must pay an immediate fee to the BZSt.
<b>Greed</b>	Someone convinces you that a mere \$10.00 investment will pocket you \$10,000 or more.
<b>Curiosity</b>	Someone sends you a voucher for trying a software that (in theory) is not yet on the market
<b>Helpfulness</b>	Playing on the basic desire of humans to trust and help one another – collecting charity and donations for a false cause
<b>Urgency</b>	You receive an email from a vendor you use indicating that they need to confirm your credit card information ASAP

How to reach a particular emotional state? Through **manipulation**

# Manipulation Tactics

Social engineers often draw on one or several **compliance techniques** to effectively **manipulate** their victims:

1. Friendship or liking: People comply easier when the request comes from a **friend** or someone they like. Social engineers will seek common ground and establish a friendship to get the target to comply with their request.
2. Commitment or consistency: . Once the target has complied with the first request, they are much more likely to agree to the rest. In social engineering, this could mean asking for a simple, easy thing first, and then **slowly continuing** with more detailed and personal requests.
3. Scarcity: People are more likely to agree to a request if they feel the **offer is scarce** or will only be available for a short period of time. Social engineering uses this technique to use the target's fear of missing out against them.

# Manipulation Tactics

4. Reciprocity: People are likelier to comply with a request if they have been treated well by the person making the request. For example, the social engineer could have done the target a **small favor**, in order to use their need for reciprocity against them.
5. Social validation: People are more likely to comply with a request if they consider it the socially correct thing to do. The social engineering attack could be framed as a **socially-expected request**, such as participating in a donation or joint effort.
6. Authority: Many people are especially trusting towards official authorities inside of an organization such as IT Support, Management, or Security. If a social engineer **camouflages as an authority** or a legitimate entity, the target is more likely to comply with the request.

**Which method works better?**

# Myers-Briggs Type Indicator (MBTI)

Not all individuals are susceptible to the same attack, but instead each of us is likely to **succumb** to a different type of **manipulation tactic**.

⇒ *Different personalities will be susceptible to different types of tactics.*

The MBTI is a preference model that defines **16 personality types** derived from **4 dimensions**, each of which is a dichotomy:

- **Extroversion-Introversion:** Refers to the way people focus their attention.
- **Sensing-Intuition:** Relates to the way people gathers information.
- **Thinking-Feeling:** Intends to show how people primarily make judgments.
- **Judging-Perceiving:** How people interact in general with the outer world.

From each dimension, a person can have **one of the either-or** characteristics.

# Myers-Briggs Type Indicator (MBTI)

<p><b>E/I</b></p> <p><b>MBTI: Extrovert/Introvert is the way we prefer to focus attention.</b> The <b><u>(E)extrovert</u></b> is interested in the external environment, i.e. people and objects that are outside the individual. Talk through problems. The <b><u>(I)ntrovert</u></b> - interested in the world of concepts and ideas, the inner world. Reflects before acting.</p>	<p><b>S/N</b></p> <p><b>MBTI: Sensing/iNtuition - the way we gather information.</b> <b><u>(S)ensing</u></b> - rely on their senses to gather information from the outside world. Trust experience, focus on what is real and factual. <b><u>i(N)tuition</u></b> - rely on hunches and own thought processes to gather information. Like ambiguity, enjoy thinking about the future.</p>
<p><b>T/F</b></p> <p><b>MBTI: Thinking/Feeling relates to the way we make judgments and decisions.</b> <b><u>(T)hinking</u></b> - analyses facts objectively and makes decisions based on cause and effect. Objective logic. <b><u>(F)eeling</u></b> – subjective decision making drawing conclusions based on empathy with the views of others (their heart to rule their head). Common ground and harmony with others.</p>	<p><b>J/P</b></p> <p><b>MBTI: Judging/Perceiving describes the way people like to live their lives, either by gathering information or drawing conclusions.</b> <b><u>(J)udging</u></b> - prefer to live in a structured, systematic, planned and organised way. Enjoy decision making and planning. <b><u>(P)erceiving</u></b> - prefer to gather information, usually easily side-tracked by things that looks more interesting. Keeps options open. Enjoys last minute time pressure.</p>



# Mapping Tactics to MBPI

**(E)xtrovert/(I)ntrovert**

**Extrovert** → Liking/Similarity

**Introvert** → N/A

**(S)ensing/I(N)tuition**

**Sensing** → Commitment/Consistency

**Intuition** → N/A

**(T)hinking/(F)eeling**

**Thinking** → Authority

**Feeling** → Social Proof

**(J)udging/(P)erceiving**

**Judging** → Reciprocation

**Perceiving** → Distraction

Perform **targeted training** based on the type of attack  
the individual is susceptible to 😊 → ongoing research



# Privacy in Online Social Networks (OSNs)

OSNs are the perfect gateways for social engineering practices:

- OSNs affordances can be easily leveraged to deceive other users (e.g., *anonymity, impersonation*).
- Attackers can reach within seconds a wide range of potential victims through the communication channels of OSNs.
- OSNs are spaces **where people make their private life public!**



Personal information disclosed in OSNs help attackers to create a **profile** of their potential victims.

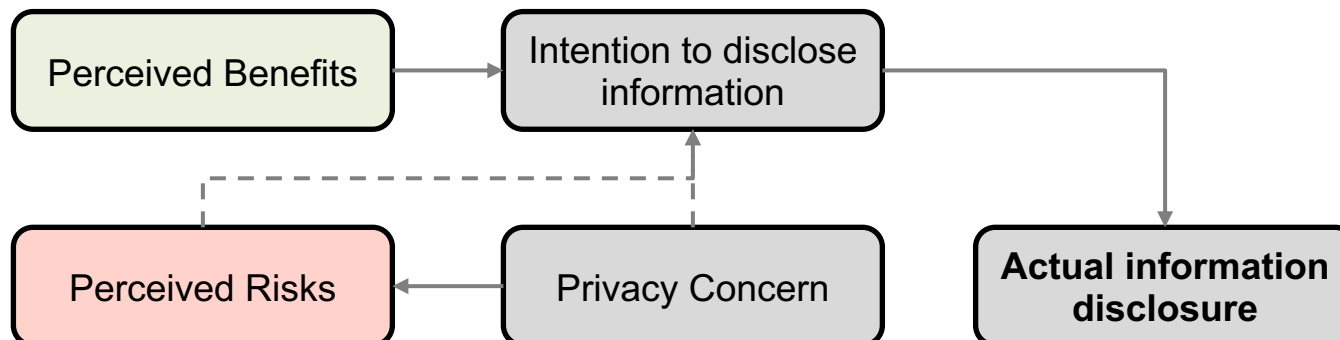
# Online Self-Disclosure

Users are not aware about the **risks** of unrestrained self-disclosure practices in OSNs (e.g., social engineering, harassment, etc.).

**X** Problem: Social media platforms lack **risk cues** inside both, their layouts and privacy policies!

**Privacy calculus**: Performing a (rational) assessment of the risk and benefits linked to personal information disclosure.

**X** Problem: Privacy decisions are mostly driven by **cognitive heuristics** instead of **rational risk estimations**.



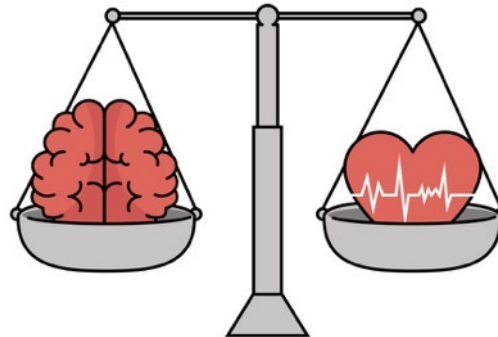
# Cognitive Biases and Heuristics

**Heuristics** (or rules of thumb) are short-cuts in decision making:

⇒ Individuals use heuristics when bounded rationality prevents the exploration of all possible outcomes.

**Cognitive Biases:** Systematic errors in judgements and behaviors:

⇒ They do not necessarily imply odd or “wrong” behavior (they are deviations from rational choices).



Biases are the **resulting gaps** between normative behavior and the heuristically determined behavior.

# Cognitive Biases and Heuristics

## The Bandwagon Effect (anchoring or social compliance):

- When deciding what to post on OSNs, one may be vastly affected by what others post, and set that as an anchor.
- People tend to take the example of their trusted peers as a reference point for what is appropriate to post and emulate them.



# Cognitive Biases and Heuristics

## Expectancy violation:

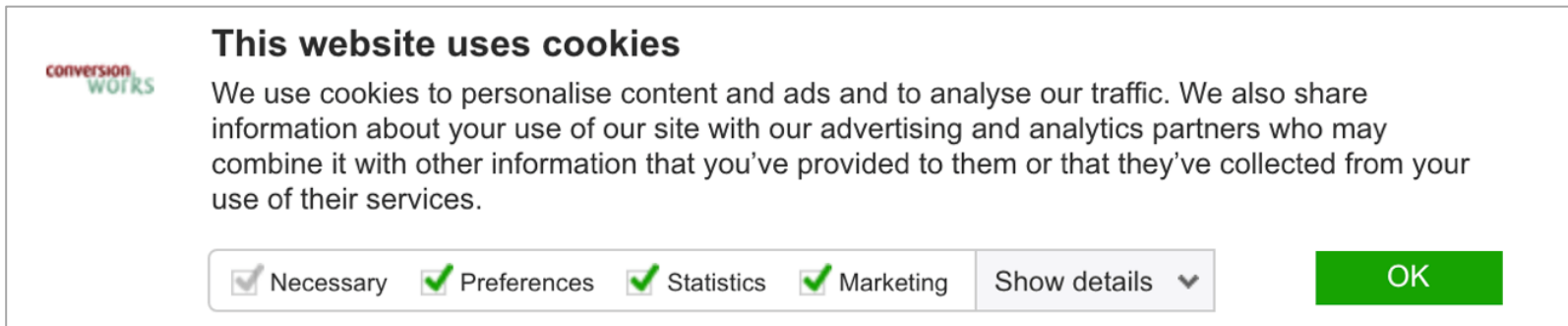
- Consists of diminishing the amount of self-disclosure if the credibility of the platform is perceived as low (e.g., <http://thebiguglywebsite.com>)
- Graphical interfaces can have a large **credibility impact**.



# Cognitive Biases and Heuristics

## Status Quo:

- Refers to individuals' affinity for default choices.
- Users usually assume that the default configurations of privacy tools protect them, without reviewing the settings.



**conversion works**

### This website uses cookies

We use cookies to personalise content and ads and to analyse our traffic. We also share information about your use of our site with our advertising and analytics partners who may combine it with other information that you've provided to them or that they've collected from your use of their services.

Necessary  Preferences  Statistics  Marketing  Show details ▼

# Cognitive Biases and Heuristics

Heuristics can be “**positive**” or “**negative**” depending whether they promote information disclosure or not:

- ⇒ *Social compliance* is a **positive** heuristic, whereas *expectancy violation* is a **negative** one.
- ⇒ Cognitive heuristics are mainly triggered by cues.



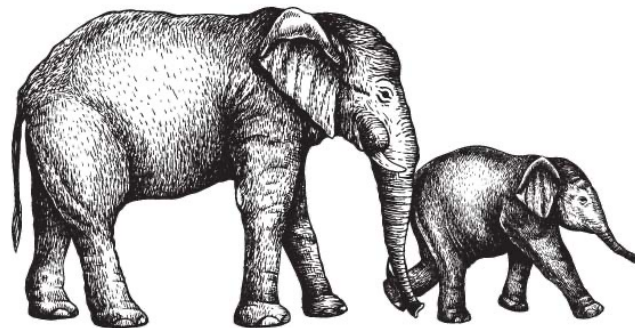
OSNs render cues that mainly trigger **positive heuristics!!!** ⇒ People share their data despite the consequences

# Privacy Nudges

When aiming to improve people's cybersecurity choices, we must consider that users are subjected to different cognitive biases:

- ✗ Biases need to be **mitigated** to prevent unintended outcomes.
- ✓ Biases can be **leveraged** to encourage beneficial behavior.

Nudges: Introduction of **small changes** in a **choice architecture** with the purpose of **encouraging** (persuade) a certain user behavior.



Scholars have elaborated on several **nudging solutions** to support users' privacy and security decisions inside and outside OSNs.



# Privacy Nudges: Examples

Information provision aims to counteract the negative effects of *availability* and *overconfidence* biases:

- Overconfidence: Underestimation of the chances that one might be subject to a negative event.
- Availability: Influence of salient cues that may not be effective signals of possible adverse events.

**My Account password**

Security level of this password: Not very secure

At least 6 alphanumeric characters.

Show password

---

**Password**

Strong

Show password

Include capital letters, special characters, and numbers to increase your password strength

Personalize your deals

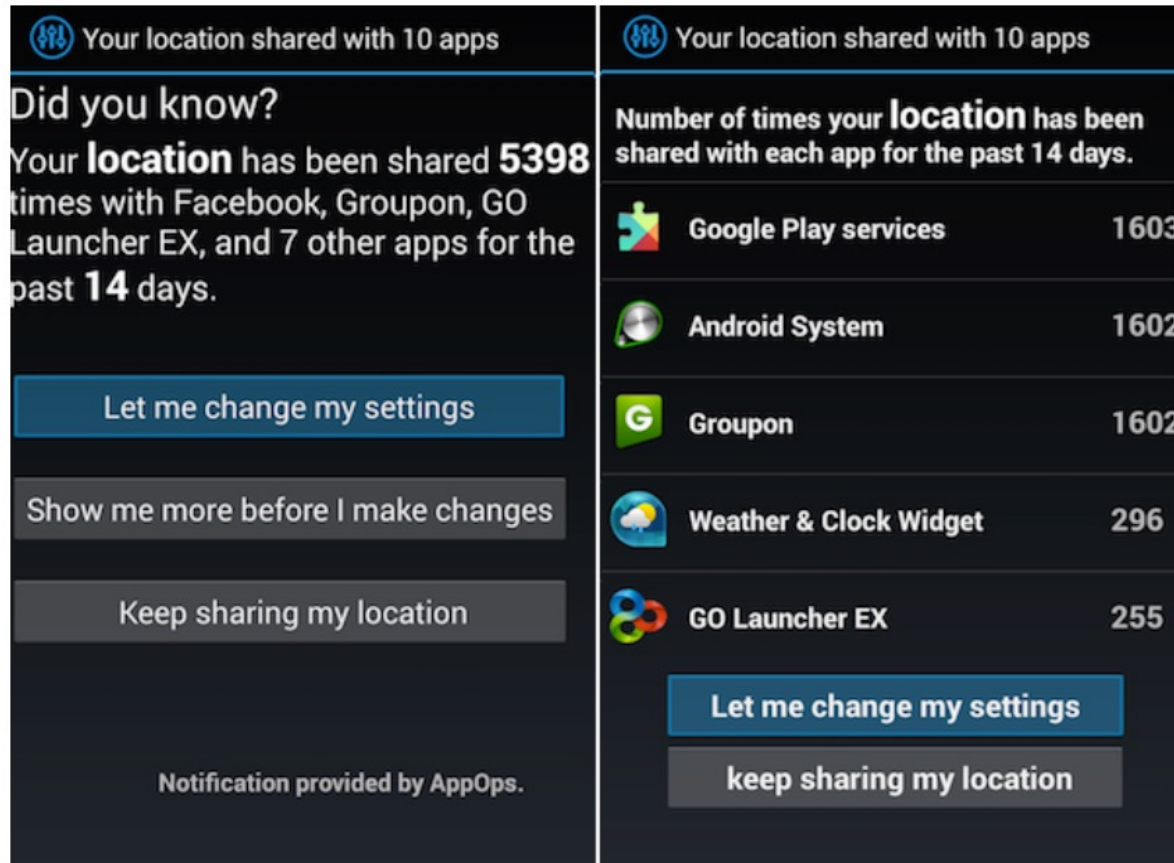
Book faster

# Privacy Nudges: Examples



WiFi scanners aim to encourage **secure networks selection**.

# Privacy Nudges: Examples



Privacy nudge for **location sharing control** in Android apps



# Privacy Nudges: Examples

The screenshot shows a social media post creation interface. At the top, there are three options: "Update Status", "Add Photo / Video", and "Ask Question". Below these is a text input field containing the text "heat in the moment". Underneath the text field is a row of icons for adding people, location, and a dropdown menu currently set to "Friends". A blue "Post" button is on the right. A yellow banner at the bottom of the interface contains the text: "Your post will be published in 3 seconds. Post Now | Edit It | Cancel".

Nudge for promoting **safer textual publications** in OSNs

# Nudges v.s. Recommender Systems

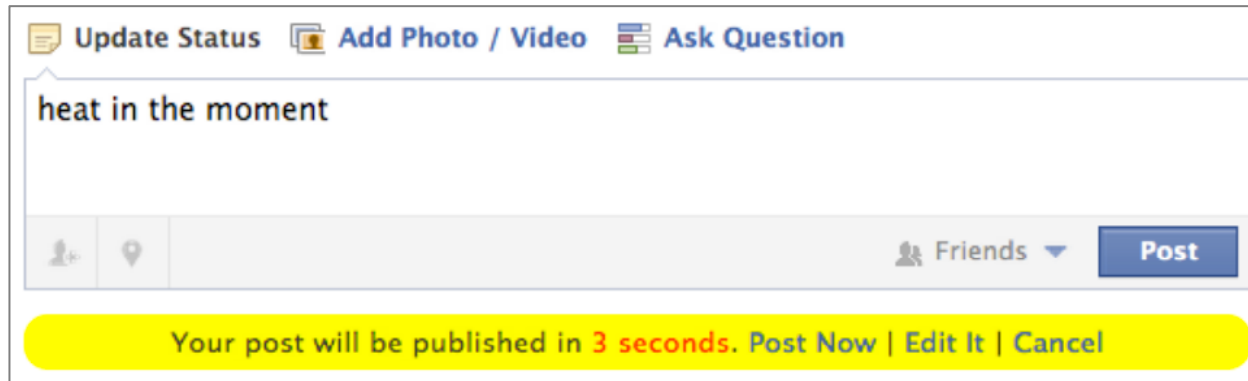
**Recommender systems** provide suggestions for items that are most likely of interest to a particular user.

- ⇒ Suggest items that are within the user's current interest area.
- ⇒ Examples: Netflix, Amazon, YouTube...

**Nudges** aim to provide recommendations that, in some respects, are outside the users' primary interests or requirements:

- ⇒ The nudging goal might not match the original interests or requirements of the user.
- ⇒ Nudges are rather about making the user stretch, to achieve something in line with the nudging goal.
- ⇒ The goal is to change users' behavior for the **common good**.

# Issues and Improvement Areas



- ✗ The purpose of the intervention is not completely clear.
- ✗ The same warning message is shown to all the users.
- ✗ No countermeasure or protective action is recommended.

When possible, nudging solutions should:

Target individuals' reflective reasoning ⇒ **risk cues!**

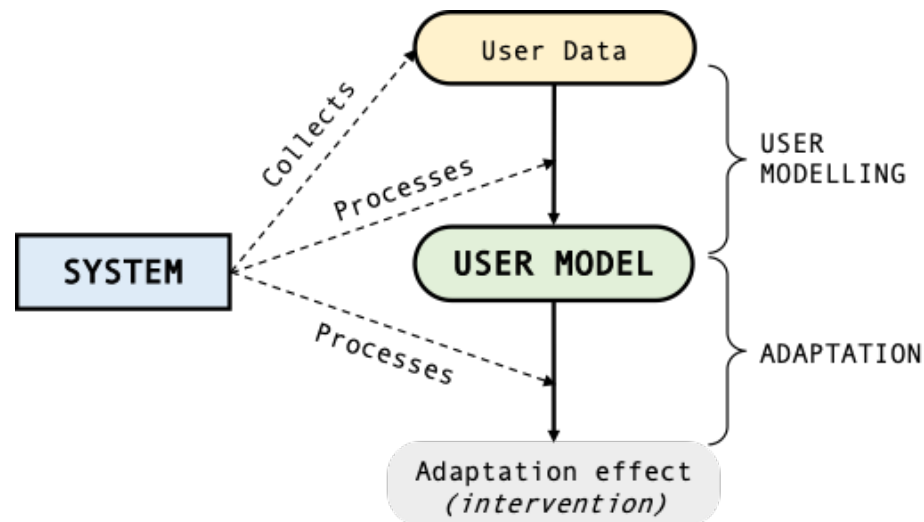
Adapt to each user's goals/expectations ⇒ **personalization!**

Recommend coping mechanisms ⇒ **audience management!**

# Privacy Nudges

Personalized nudging solutions employ **Artificial Intelligence (AI)** to understand and anticipate the (privacy) needs of each user:

- User Model: A set of adaptation variables that guide the personalization of behavioral interventions (e.g., *privacy attitudes*).

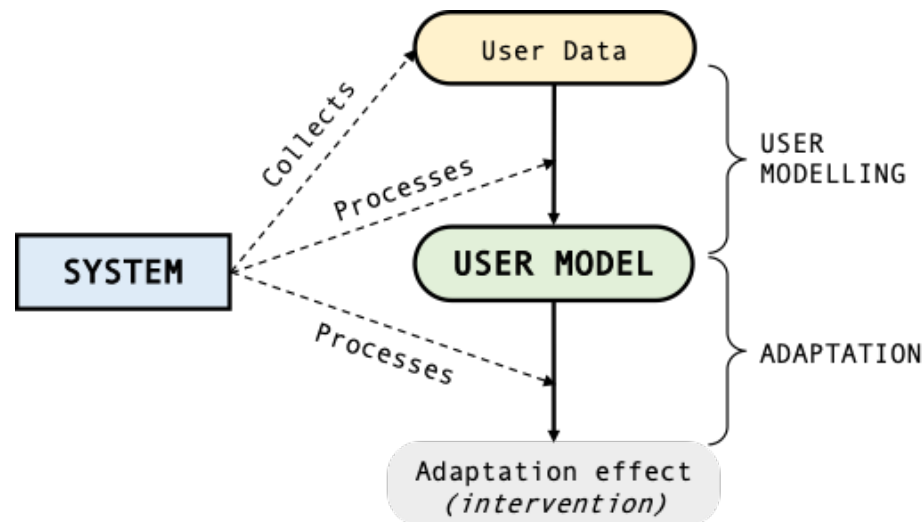


User models can be generated either **explicitly** (e.g., set-up questionnaire) or **implicitly** (e.g., behavioral data)

# Privacy Nudges

Personalized nudging solutions employ **Artificial Intelligence (AI)** to understand and anticipate the (privacy) needs of each user:

- User Model: A set of adaptation variables that guide the personalization of behavioral interventions (e.g., *privacy attitudes*).



User models can be generated either **explicitly** (e.g., set-up questionnaire) or **implicitly** (e.g., behavioral data)



# Privacy Attitudes (Westin)



People's privacy attitudes can be identified using a questionnaire:

- **Unconcerned** users are the less privacy protective.
- **Fundamentalists** seek actively for privacy and data protection.
- **Pragmatists** are in an intermediate position.

Example: Fundamentalists should only be informed on very-high privacy risks, whereas pragmatists also about low risks.

# Privacy Attitudes (Westin)

Indicate how much do you agree/disagree with the following statements:

- **Q1:** “Consumers have lost all control over how personal information is collected and used by social media platforms”.
- **Q2:** “Most platforms handle the personal information they collect about consumers in a proper and confidential way”.
- **Q3:** “Existing laws and software development practices provide a reasonable level of protection for consumer privacy today”.

Answering options: *strongly agree, somewhat agree, somewhat disagree, strongly disagree, don't know*

- **Fundamentalists** agree (strongly or somewhat) to Q1 and disagree (strongly or somewhat) to Q2 and Q3.
- **Unconcerned** disagree (strongly or somewhat) with Q1 and agree (strongly or somewhat) with Q2 and Q3.
- **Pragmatists** are those with any other pattern of responses.

# Multiparty Privacy Conflicts

Overall, current preventative nudges focus on **individual** self-disclosure risks:

- ✗ They do not consider unwanted incidents that may occur when sharing content that also compromises the **privacy of others**.

Situations in which personal information of others is unintentionally exposed to the public are frequent:

- ☹ People **sharing pictures** of their friends **without consent**.
- ☹ People **tagging others** in publications without taking their individual privacy preferences into account.

⇒ **Multiparty Privacy (MP)** takes a **collective view** on the norms and boundaries of information disclosure.

MP elaborates on the **conflicting privacy preferences** among the **co-owners** of particular data items.

# Methods and Strategies

Overall, current **methods and strategies** for counteracting MP conflicts in OSNs can be classified into:


- **Dissuasive:** *“...aim to make uploaders reflect on the implications of sharing a given item and raise awareness about the consequences of unilateral decisions”.*
- **Precautionary:** *“...automate collaborative practices and force uploaders to collaborate with data subjects or otherwise limit the shared content”.*

Precautionary mechanisms can be further divided into:

- Audience modification: Mechanisms that modify an item’s audience (e.g., who can see a photo).
- Item modification: Mechanisms that obfuscate the item to be shared (e.g., blurring faces in a photo).

# Precautionary: Item Modification

**Photo Privacy Setting** ✕


 **Privacy**

We identify the following content in your photo which could cause privacy issues:

- Beer can**
- Bob's face**
- Drew's face**
- Unknown person's face**

For the selected content, which obfuscation would you like to apply?

**Inpainting** ▾



Cancel Save



# Precautionary: Item Modification

## Secured Online Social Network

Adam Ebert

Home Disclosure Detection Login Photos Chat Room Forum

**Message:** My hometown is **Tokyo**. My favorite food is sushi. After graduating from Tokyo University, I studied at **Harvard University** for three years as a computer science major.

Families **Tokyo - Harvard University** ▾

Disclose  
 Not disclose  
Edit message  
Change Synonyms

Bob Smith

My hometown is **Tokyo** ▾. My favorite food is sushi. After graduating from Tokyo University, I studied at **Harvard** ▾ for 3 years as a computer science major.

Old friends **Tokyo - USA** ▾

Disclose  
 Not disclose

Dave Henderson

---

Disclose  
 Not disclose  
Edit Message  
Change Synonyms

Ellen Anderson

My hometown is **Tokyo** ▾. My **favourite** food is sushi. After graduating from Tokyo University, I studied at **USA** ▾ for three years as a **USA** ▾ ce major.

Post

- USA
- US
- U.S.A
- U.S.
- United States

# Dissuasive

**Warning**

You are about to share a picture featuring several individuals. Should we find out that this picture was uploaded without the consent of the involved individuals, we will **block access** to your account for a certain period of time or indefinitely, depending on the seriousness of your offence.

**CANCEL**      **CONTINUE**

(a) Account Locked Strategy (AL).

**Warning**

You are about to share a picture featuring several individuals. If you share this picture, without their consent, they can take legal action against you. Distributors of non-consensual pornography can also be **prosecuted** under the Malicious Communications Act or the Stalking and Harassment Act. Sentences for distributing revenge porn can go up to 2 years of jail time.

**CANCEL**      **CONTINUE**

(b) Law Threat Strategy (LT).

**Warning**

You are about to share a picture featuring several individuals. Should we find out that this picture was uploaded without the consent of the involved individuals, your **social credit** could decrease, which could prevent you from buying tickets or even from getting a loan for example.

**CANCEL**      **CONTINUE**

(c) Social Score Strategy (SS).

**Warning**

You are about to share a picture featuring several individuals. Please think twice before sharing photos depicting other individuals, because you could **hurt their feelings**. Don't do to others what you wouldn't want done to yourself.

**CANCEL**      **CONTINUE**

(d) Empathy Strategy (E).

# Ethical Challenges

As personalization in nudges increases, so do **concerns** related to *transparency, fairness, explainability, algorithmic biases*.

⇒ Inherited from the underlying principles of AI technologies!

User models and adaptation mechanisms should be **scrutable** for preventing *inaccurate, unfair, biased, or discriminatory* interventions.

There are also challenges related to the impact on people's individual and collective behavior:

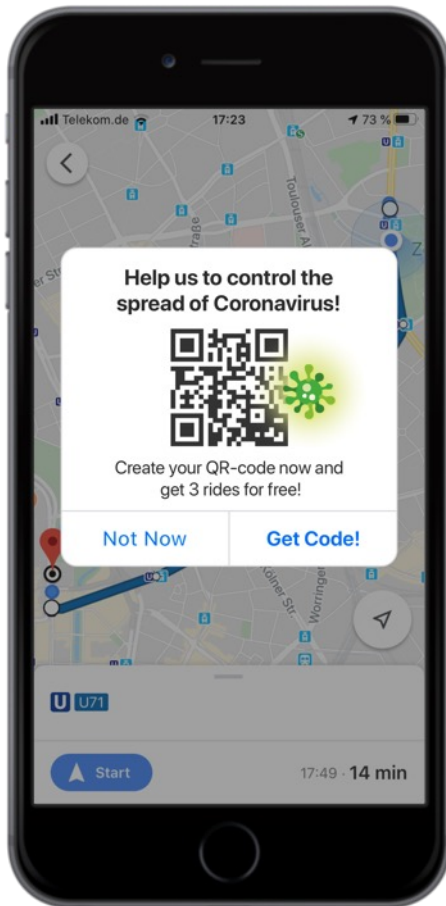
- Nudges may not necessarily contribute to users' welfare.
- Could even be used for questionable and unethical purposes.

A fine line between **persuasion, manipulation** and **coercion**!



# Ethical Challenges

Example: Nudge to incentive the use of COVID-19 tracing mechanisms.



The **argument**:

- ✓ Encourage people to provide their location and body temperature on behalf of public safety.

The **real purpose**:

- ✗ Another attempt to increase mass surveillance.

Ethical Questions:

- 🤔 *Who should benefit from nudges?*
- 🤔 *Should users be always informed about the presence of a nudge?*
- 🤔 *How nudges should (not) influence the users?*

# Ethical Guidelines

Persuasive means target primarily people's automatic and **subconscious** processing system:

⇒ This can compromise users' *agency* and *autonomy* since they may **not be aware** of the presence of a nudge.

**Check-lists** can be employed to verify whether a nudging solution comply with principles of justice, beneficence, and respect:

- ✓ To preserve user's autonomy, we must ensure that all the **original options** of a choice architecture are made **available**.
- ✓ Users should always be nudged towards behaviors that **maximize their welfare** rather than the interests of others.

Nudges should target, when possible, individuals' **reflective reasoning** (e.g., through risk cues) to avoid potential manipulation effects.



# Questions ?

