UNIVERSITY OF TARTU
Faculty of Science and Technology
Institute of Computer Science
Computer Science Curriculum

Beatriz Pontes da Costa Reis

An Approach for Designing Responsible Privacy Heuristics

Master's Thesis (30 ECTS)

Supervisor(s): Mohamad Gharib, PhD

Tartu 2025

An Approach for Designing Responsible Privacy Heuristics

Abstract:
Privacy compliance is a major concern for legal entities handling Personal Information (PI), as noncompliance leads to substantial fines. Regulations require these entities to implement privacy protection mechanisms (privacy solutions) and inform data subjects (DSs) about PI processing. However, DSs often struggle to understand relevant information and effectively use these mechanisms, leaving their privacy vulnerable. Privacy heuristics (PHs) offer a potential solution by assisting users in making informed decisions. Yet, their design is complex, prone to bias, and, if done irresponsibly, can lead to unethical or manipulative outcomes.

This thesis addresses these challenges by developing an approach that provides design principles for guiding and evaluating Responsible Privacy Heuristics (RPHs) in privacy-aware systems. Following the Design Science Research methodology, we formulated the principles to satisfy six meta-requirements derived from ethical principles: Integrity, Non-manipulation, Beneficence and Non-maleficence, Autonomy and Control, Context-aware and Accessible, and Regulatory Compliance. Each principle is paired with acceptance criteria that practitioners can use to verify correct application. The clarity and applicability of the resulting eleven design principles, as well as the validity of their acceptance criteria, were evaluated by privacy domain experts.

We demonstrate the applicability of the approach through a practical example, following the steps of the methodological process. The resulting design was validated via a moderated A/B test with 12 end-users. Participants were asked to complete demographic questions, read a scenario, interact with their assigned design version, and then respond to a post-task questionnaire that assessed perceived usability, perceived informed decision, perceived autonomy, and perceived consequences awareness. In addition, we evaluated informed decision and decision-awareness to measure the new privacy solution's effectiveness. The results show that the RPH version matched the standard PH version in usability, while being slightly more effective in preventing the selection of privacy-invasive options and enabling informed decision-making, without compromising user autonomy.

Keywords:
Usable privacy; Responsible privacy heuristics; Design principles; Ethical design

CERCS:
P170, Computer science, numerical analysis, systems, control

Vastutustundliku privaatsuse heuristika kujundamise lähenemisviis

Lühikokkuvõte:

Privaatsusnõuete järgimine on isikuandmeid (PI) käitlevate juriidiliste isikute jaoks suur murekoht, kuna mittevastavus toob kaasa märkimisväärseid trahve. Määrused nõuavad, et need üksused rakendaksid privaatsuse kaitsemehhanisme (privaatsuslahendusi) ja teavitaksid andmesubjekte (DS-e) PI töötlemisest. DS-idel on aga sageli raskusi asjakohase teabe mõistmise ja nende mehhanismide tõhusa kasutamisega, mis muudab nende privaatsuse haavatavaks. Privaatsusheuristikad (PH-d) pakuvad potentsiaalset lahendust, aidates kasutajatel teha teadlikke otsuseid. Ometi on nende ülesehitus keeruline, altid eelarvamustele ja vastutustundetu tegutsemise korral võivad need viia ebaeetiliste või manipuleerivate tulemusteni.

See väitekiri käsitleb neid väljakutseid, töötades välja lähenemisviisi, mis pakub disainipõhimõtteid vastutustundliku privaatsuse heuristika (RPH) juhtimiseks ja hindamiseks privaatsusteadlikes süsteemides. Disainiteaduse uurimistöö metoodikat järgides sõnastasime põhimõtted, mis vastavad kuuele eetilistest põhimõtetest tulenevale metanõudele: terviklikkus, mittemanipuleerimine, heategevus ja mittekahju, autonoomia ja kontroll, kontekstiteadlikkus ja ligipääsetavus ning regulatiivne vastavus. Iga põhimõttega on seotud vastuvõtukriteeriumid, mida praktikud saavad kasutada õige rakendamise kontrollimiseks. Saadud üheteistkümne disainipõhimõtte selgust ja rakendatavust ning nende vastuvõtukriteeriumide kehtivust hindasid privaatsusvaldkonna eksperdid.

Me demonstreerime lähenemisviisi rakendatavust praktilise näite abil, järgides metodoloogilise protsessi samme. Saadud disaini valideeriti modereeritud A/B-testi abil 12 lõppkasutajaga. Osalejatel paluti vastata demograafilistele küsimustele, lugeda stsenaariumi, suhelda neile määratud disainiversiooniga ja seejärel vastata ülesandejärgsele küsimustikule, mis hindas tajutavat kasutatavust, tajutavat teadlikku otsust, tajutavat autonoomiat ja tajutavate tagajärgede teadlikkust. Lisaks hindasime teadlikku otsust ja otsusteadlikkust, et mõõta uue privaatsuslahenduse tõhusust. Tulemused näitavad, et RPH versioon vastas kasutatavuse poolest standardsele PH versioonile, olles samal ajal veidi tõhusam privaatsust riivavate valikute valimise ennetamisel ja teadliku otsuste tegemise võimaldamisel, ilma et see kahjustaks kasutaja autonoomiat.

Võtmesõnad:

Kasutatav privaatsus; Vastutustundliku privaatsuse heuristika; Disainipõhimõtted; Eetiline disain

CERCS:

P170, Computer science, numerical analysis, systems, control

Contents

1 Introduction

1.1 Problem Statement

In 2009, European Commissioner for Consumer Protection Meglena Kuneva stated that *"Personal data is the new oil of the Internet and the new currency of the digital world"* [19]. Spiekermann et al. [33] highlighted the growing challenges of personal data (PD) (also called Personal Information (PI)) markets and privacy, emphasizing how PD has evolved into a valuable asset for both companies and consumers [9]. They also discuss its role in powering a wide range of services and products, including personalized advertisements, recommendation systems, consumer risk analysis, and even as a product itself—particularly in user-generated content on social networks.

As personal data increasingly becomes a product and service, integrating privacy into digital systems has become essential [27]. Regulations such as the General Data Protection Regulation (GDPR) enacted in the EU [8], Brazil's General Personal Data Protection Law (LGPD) [6] and Japan's Act on the Protection of Personal Information (APPI) [15], among others, impose legal obligations to safeguard data subjects (DS) and their privacy. These laws aim to protect the DS and prevent the mismanagement of their PI, including misuse, excessive processing, improper storage, and unauthorized third-party sharing [11].

Although legal entities handling PI are required to provide DSs with privacy protection mechanisms and disclose how their PI will be processed, the responsibility of understanding this information and effectively using these mechanisms still falls on DSs [16]. This poses a challenge, as users have varying levels of digital literacy and may struggle with the legal jargon found in privacy policies or cues.

This challenge reflects a broader issue concerning the need to consider social and human factors when designing and managing processes involving PI. Business processes, traditionally, focus on efficiency and compliance, but when it comes to privacy, they must also account for the psychological and behavioral dimensions of user interaction. Users are not merely endpoints in a process; they are active participants with diverse expectations, preferences, and vulnerabilities. Failing to address these aspects can lead to disengagement, mistrust, or even harm.

A promising solution is the use of heuristics—popularly known as "mental shortcuts" [14]—more specifically privacy heuristics that can help users make informed decisions and take appropriate actions [10]. However, designing privacy heuristics is complex and prone to bias [14]. If not designed responsibly, they may influence the DS judgments or decisions in a manner that is considered unethical, immoral, or socially irresponsible.

This thesis aims to address and mitigate this burden by developing an approach that offers design principles to guide the design and evaluation of Responsible Privacy Heuristics (RPHs) for privacy-aware solutions. These principles aim to empower users, uphold their autonomy and self-determination, and facilitate informed decision-making.

1.2 Objectives of our research

The goal of this thesis is to reduce the burden DSs face when engaging with privacy mechanisms and making decisions about their PI. To address this, we propose an approach that introduces a set of design principles aimed at guiding the creation and evaluation of Responsible Privacy Heuristics (RPHs). These principles are intended to be used by designers and developers involved in crafting privacy-related interfaces—such as privacy policies, settings, and cues. To accomplish this objective, we are committed to answering three research questions:

- **RQ1. What are RPHs?**
  We answer RQ1 in Section 2, when discussing existing work about ethical design and privacy heuristics.
- **RQ2. How can we design RPHs for various privacy solutions?**
  In Section 3, we outline the methodology adopted to conduct this research, the process to create the approach and the methodological process designers must follow to use the approach, consequently addressing RQ2.
- **RQ3. How can we validate the usability and effectiveness of the developed RPHs in the final privacy solution?**
  We respond to RQ3 in Section 5, when validating the privacy solution derived from the application of the approach.

1.3 Contributions of the Thesis

The main contribution of this thesis is an approach for designing and developing RPHs. This requires achieving the following three sub-objectives:

- A clear definition of RPHs;
- An approach to develop and evaluate RPHs, which offers design principles and a systematic process to be followed by designers and software engineers while using the approach;
- A demonstration of the usability and effectiveness of the approach in aiding software engineers in designing and developing RPHs.

1.4 Structure of the Thesis

This thesis is structured into seven sections. Section 1 introduces the research topic and problem statement, along with the thesis objectives, which outlines the overall goal and research questions. Section 2 provides a literature review covering key concepts related

to privacy heuristics and ethics in decision-support mechanisms. Then, develops the definition of responsible privacy heuristics, and briefly reviews prior solutions related to ethical decision-support mechanisms, pointing out their similarities and differences in regards to this work. Section 3 outlines the methodology used to develop the approach, details the process followed to create the design principles, and presents the methodological process that designers should follow to derive RPHs. Section 4 illustrates the first three steps of this process by demonstrating how the design principles can be applied in practice, starting with an initial design based on existing privacy heuristics, followed by their refinement, and concluding with a redesign using the proposed RPHs. Section 5 presents the validation of the design principles by experts, and the final step of the methodology: validation of the RPHs solution, which was carried out through an A/B test experiment involving 12 participants and discussion of the results. After the result, in Section 6, we address the threats to validity, and finally, Section 7 summarizes the main findings and contributions, and directions for future work.

2 Baseline

2.1 Privacy Heuristics

Heuristics are often described as "rules of thumb" or "mental shortcuts" that can be used to speed up the decision-making process [13, 14]. While broadly defined, they are commonly seen as problem-solving methods that do not guarantee optimal solutions [14]. They help with both ill-defined and well-defined problems by reducing cognitive effort [10]. Heuristics can be instinctive (automatic) or deliberate, with experience transforming one into the other over time [14].

Heuristics play a key role in online privacy, especially in PI disclosure, where they are termed Privacy Heuristics (PHs). Sundar et al. [34] classify PHs into three different types of context: Social (influence of group dynamics in disclosure), Personal (self-protection or self-reward in disclosure), and Technological (interface elements shaping behavior in disclosure). Vincent et al. [20] proposes six superordinate PH classes: Prominence (perceived credibility and trust), Network (social influence), Reliability (trust in professional and consistent design), Accordance (alignment with one's beliefs), Narrative (impact of storytelling) and Modality (influence of new technologies).

In short, privacy heuristics simplify privacy decisions by enabling quick, efficient choices while ignoring some information. They can be deliberate or automatic, influenced by the environment, experience, and cognitive biases. Table 1 compiles key heuristics from [34, 20, 18], which can influence privacy decisions.

Table 1. Heuristics that can influence privacy decisions.

Affect Heuristic. People judge objects or events by associating them with positive or negative feelings.
Anchoring. Under uncertainty, people tend to be biased towards a reference point, or "anchor".
Choice Overload. Too many options make people feel overwhelmed and influence their judgment negatively.
Contrast effect. People's decision is influenced by comparing one instance with another, instead of relying on impartial standards.
Framing. People's choice frame is set up in a way to manipulate/control the user's decision.
Functional Fixedness. People tend to fixate on a specific use of an object
Instant gratification. People prioritize quick rewards at the expense of future gains.
Loss Aversion. People prefer to avoid losses rather than acquire equivalent gains.
Optimism bias. People tend to underestimate the chances of experiencing negative events and overestimate positive ones.

Social Norms. People's behavior is influenced by social norms that either play a part in guiding or constraining it.
Status Quo/Default Effect. People tend to favor options that maintain the current state over those that introduce change.
Authority. Recognized brand, institution or person that vouches for the security, influences disclosure.
Bandwagon. People are influenced by the decision of many or the majority to disclose information.
Reciprocity. People are more likely to share information with someone who has disclosed theirs to them.

2.2 Responsible Privacy Heuristics

2.2.1 Definition and characteristics

To develop a clear definition of responsible privacy heuristics, it is essential to examine how ethical considerations have been applied to decision-support mechanisms. This includes analyzing both unethical and ethical design approaches, as well as their impact on users' decision-making processes. We begin by reviewing unethical design practices, particularly the use of deceptive patterns and their consequences in privacy-related decisions. We then turn to ethical approaches, including the concept of fair patterns and the responsible use of nudges as supportive mechanisms. This ethical foundation ultimately culminates in the definition of responsible privacy heuristics.

Deceptive patterns, also known as "dark patterns", were first described by Brignull in 2010 [2], as *"tricks used in websites and apps that make you do things that you didn't mean to, like buying or signing up for something"*. These unethical patterns are often characterized by their coercive, manipulative and/or exploitative nature, aiming to guide the user into decisions that primarily benefit the service provider [4], often at the expense of the user's best interest [21, 18]. These patterns exploit user biases and heuristics to trigger automatic, fast, and intuitive decision-making. They frequently alter the choice architecture by hiding or obstructing privacy-preserving options, instead promoting those that encourage greater PI disclosure [29]. This manipulation prevents users from making informed, conscious choices, potentially leading to harmful decisions regarding their personal data.

Kitkowska [18] has identified unethical patterns in the existing literature and organized them into taxonomies. Based on her work, Table 2 presents examples of Privacy Deceptive Patterns (PDPs), the psychological effects (heuristics and biases) they may trigger, and their potential impact on user decisions. This table is not intended to provide an exhaustive listing of PDPs, but rather to illustrate the concept and lay the groundwork for an ethical approach.

Table 2. Privacy Deceptive Patterns, triggered heuristics and effects on user [18].

Privacy Deceptive Patterns	**Heuristic(s)**	**Effect on user**
Privacy Zuckering: is the use of deceptive design or persuasive tactics to manipulate users into sharing more PI than they intend to.	Choice overload, Status quo and Framing.	Users share more PI than they intended and might be unaware of how their PI is being processed.
Bad defaults: user account options are predefined in a privacy-invasive manner (over-sharing), and sometimes might not even have an alternative option available.	Default effect, Status Quo, Loss aversion and Instant gratification	Same as Privacy Zuckering.
Comparison obfuscation: hinders users from easily comparing privacy policies, data collection practices, or security features across different services.	Anchoring and Optimism bias.	Users adopting privacy-invasive options
Forced action: users are forced to make choices immediately in order to use a service or product.	Instant gratification and Framing.	Users end up sharing more PI than they intended to keep using a service or product.
False necessity: persuades the user into privacy-invasive choices by claiming that the data is essential for the service.	Framing and Status Quo	Same as Forced action effects.
Just between you and us: makes false promises of confidentiality to encourage users to disclose more information.	Optimism bias and Framing.	Users might share more than they usually would due to the false sense of confidentiality.
Trick questions: deceives the user into making privacy-invasive choices with misleading or ambiguous wording.	Default effect, Framing and Anchoring.	Users will be confused and most likely misinterpret their choices.
Attention diversion: distracts the user from privacy-conscious choices by emphasizing other aspects of the UI.	Anchoring and Framing	Hinders users from properly reflecting on their privacy-related choices.

Continued on next page...

Privacy Deceptive Patterns	Heuristic(s)	Effect on user
Wrong signal: uses distinguishable icons, symbols or other elements in the UI to misguide users.	Anchoring, Framing, Affect heuristic and Authority [34]	Certain UI elements create the illusion of privacy-conscious choices, misleading users into feeling secure.
Confirmshaming: steer users to make specific choices through guilt/shame. May use UI elements to induce a certain emotional state.	Affect heuristic, Contrast effect, Default effect, Framing, Bandwagon [34]	Privacy-conservative choices might be painted as negative or selfish, manipulating users to share more data.
Last-minute consent: leverages time pressure and context to push users to consent to less optimal privacy options choices or make privacy decisions without the option to delay.	Loss aversion and Status quo	Users experience a reduced freedom of choice and might be coerced to comply with privacy-invasive choices to avoid losing progress.
Safety blackmail: users are pressured into less optimal privacy options by implying that failing to do so could result in safety or security risks.	Functional fixedness and Instant gratification.	Users end up sharing more PI than they intended, in order to enable their accounts.

While deceptive patterns have been extensively researched and existing work offers valuable insights into what should be avoided, there is a notable gap in research focusing on what should be done [4]. This thesis addresses that gap by providing a systematic approach for developing RPHs (i.e., ethical decision-support mechanisms) within the privacy context.

To counteract deceptive patterns, researchers have proposed ethical design approaches that integrate ethical principles into decision-support mechanisms. Fair patterns are considered the ethical counterparts to deceptive patterns; their aim is to empower users to make informed, voluntary decisions without manipulation or obstruction [29]. Another form of decision-support mechanism, though often debated, is the use of nudges. Nudges are typically described as a form of soft paternalism that subtly steers users toward decisions deemed to be in their best interest, without restricting their access to other options (when used ethically) [35]. However, due to their inherently paternalistic nature, which involves deliberately influencing users toward a specific option, nudges have raised ethical concerns in regards of user autonomy and freedom of choice [1, 35].

At the core of ethical design approaches lies a commitment to user empowerment, facilitating informed decision-making, and prioritizing the user's best interests [26].

Ethical design should emphasize clarity, transparency, accessibility, and simplicity to ensure that users can make truly informed choices. More specifically, they should be grounded in fundamental ethical principles, such as: Respect, Beneficence (and Non-Maleficence [17]), Justice, Integrity, and Social Responsibility [31]. Building on these principles, we define Responsible Privacy Heuristics (RPHs) as user-empowering, transparent, inclusive, and easy-to-understand decision support mechanisms that respect user autonomy and self-determination, enabling them to make informed decisions in the context of privacy.

2.2.2 Ethical approaches supporting privacy decision-making

We review the relevant literature on responsible and ethical design, with a focus on research that seeks to counteract unethical practices (e.g., deceptive patterns) through design patterns, guidelines, and frameworks. The selected studies share a common objective: to support designers in creating systems that facilitate informed decision-making while respecting and empowering user autonomy. Accordingly, this review highlights the key contributions and limitations of these works, as well as how this thesis builds upon or diverges from them.

Potel-Saville and Rocha [29] introduce a new taxonomy of deceptive patterns and their corresponding fair patterns (FPs), which aims to shift the focus from a problem-oriented towards a solution-oriented approach to deceptive pattern mitigation. They acknowledge the value of previous academic research and taxonomies, which laid a strong foundation by identifying, categorizing, and evaluating the harms and effectiveness of DPs. However, they argue that while these contributions aid in recognizing and understanding DPs, they fall short of providing practical guidance for designers seeking to counteract them. To address this gap, the authors propose a higher-level, abstract taxonomy that builds on top of previous taxonomies to ensure its applicability in multiple domains. They ground their solution on core cognitive biases that DPs exploit, making it capable of accommodating new forms of DPs, something they describe as "future proof". Their solution is displayed in Table 3.

Table 3. Usable taxonomy of dark patterns with corresponding fair patterns by [29].

Deceptive Pattern	Fair Pattern
Harmful Default: Default settings are against the user's interest.	**Protective Default**: Defaults prioritize user privacy and well-being, aligning with positive societal outcomes.
Missing Information: Selective disclosure of information.	**Adequate Information**: Users receive clear, sufficient, and relevant information without unnecessary overload.

Maze: User path to information, preferences, or choices is made unnecessarily complex.	**Seamless Path**: User path to information, preferences, or choices is as easy when they are in the user's interest as when they are in the service provider's interest.
Push & Pressure: Emotional, social, or time-based triggers pressure user decisions.	**Non intrusive information**: No manipulative nudges unless they serve user or societal benefits.
Misleading or Obstructing Language: Language is confusing, manipulative, or impedes user understanding.	**Plain and Empowering Language**: Clear, accessible, and jargon-free wording helps users make informed decisions.
More than intended: Users are led through a series of steps that force them to do or give more than they originally intended.	**Free action**: Users are empowered to understand the consequences of their choices, especially regarding spending or data sharing, without unnecessary information overload.
Distorted UX: The visual interface is designed to mislead or trap users.	**Fair UX**: The interface respects user intentions by ensuring clarity in placement, shape, size, and prominence of buttons and icons.

The fair patterns proposed by Potel-Saville and Rocha [29] share several similarities with the approach presented in this thesis: both are solution-oriented and abstract approaches designed to help designers mitigate deceptive patterns. However, this thesis distinguishes itself by placing particular emphasis on producing a solution grounded in ethical principles and focused on the privacy domain. Moreover, this thesis demonstrates the application of the solution to be followed by practitioners, including an experiment to validate it. This is missing from [29], although mentioned as future work.

Another relevant approach is Mildner's [23] Responsible Design Triangle, a user-centered framework that establishes the relationship between the user, design, and guidelines. The angles are defined as follows:

- **Design** represents the intended purpose of a system, made perceivable to users as a solution to a specific problem. It sets realistic expectations and communicates the consequences of user interactions in a transparent manner;
- **Users** engage with (digital) technologies through design, guided by their perceptions, expectations, goals, and actions;
- **Guidelines** are rules or instructions aimed at achieving particular goals. They should inform the design process through thoughtful and responsible practices that enable designers to foster ethical interactions.

Figure 1 presents the Responsible Design Triangle as a diagram situated positioned within a three-dimensional space, highlighting the contrast between ethical and unethical relationships among its angles. When ethically aligned, these components collectively inform ethical good design. Mildner [23] defines *good design* as a design that *"...utilises peoples' cognition and perception to effectively alter their choice architecture to convey a planned goal, which is easily carried out, while implications are transparent and informed decision-making is empowered."*. Conversely, when aligned with the unethical side, it results in a deceptive pattern, that manipulates and exploits the user for the benefit of service providers.

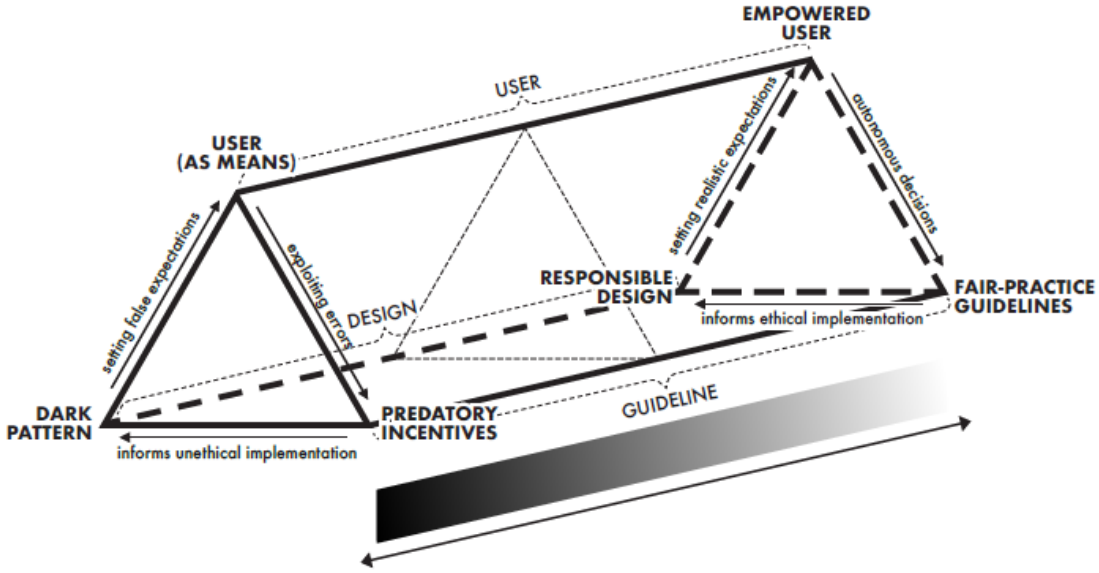


Figure 1. Diagram of Mildner's Responsible Design Triangle [23].

Our solution aligns with Mildner's Responsible Design Triangle by providing *guidelines*, in the form of design principles and a methodological process, that support practitioners in deriving responsible privacy heuristics. These heuristics are intended to guide the creation of ethical and usable designs that empower users to make informed decisions. In this regard, Mildner's framework offers a valuable, high-level perspective that complements our work. It serves as an abstract model for understanding the ethical dimensions of design. However, unlike our approach, it does not function as a hands-on, practitioner-oriented guide for the design process. This thesis addresses that gap by offering concrete, actionable steps for implementing ethical design in the privacy domain.

3 An Approach for Designing Responsible Privacy Heuristics

The approach has been developed following the Design Science Research (DSR) methodology [28], which aims to *"improve the state of practice and contribute to design knowledge through the systematic construction of useful artifacts"* [32]. In this context, an artifact refers to a form of design knowledge, such as models, methods, frameworks, guidelines, patterns [25] and many others, including design principles. DSR not only focuses on the creation of such artifacts but also emphasizes their demonstration, evaluation, and communication. Specifically, our methodology aligns with DSR's key steps while further refining some into sub-steps as illustrated in Figure 2, and described as follows:

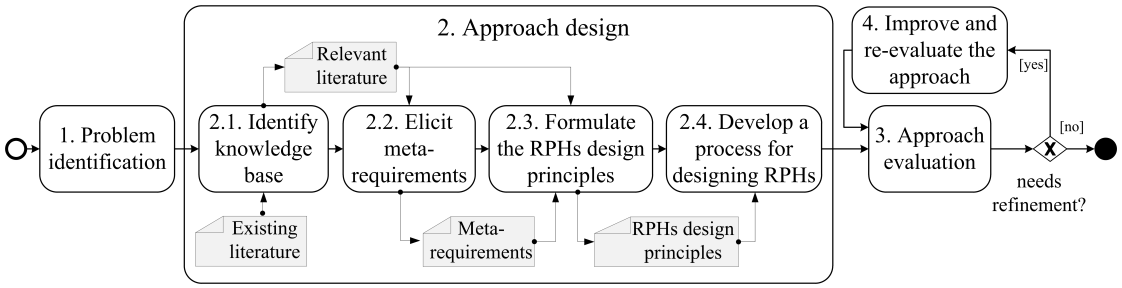


Figure 2. The process for constructing and evaluating the approach.

1. Problem identification: as discussed earlier, there is a need for an approach to guide the design and evaluation of RPHs for privacy-aware solutions.

2. Approach design: is composed of four sub-steps; *2.1. Identity knowledge base*, *2.2. Elicit Meta-requirements*, *2.3. Formulate the RPHs design principles*, and *2.4. Develop a methodological process for designing RPHs*. The first three sub-steps have been adopted following the method for developing design principles in [24], and the last step offers a systematic process for using the approach. We describe these steps in the following subsection.

3. Approach evaluation: aims at evaluating the approach based on how well it supports solutions in the problem space. In particular, we will demonstrate its applicability following the methodological process, we get experts to evaluate the design principles and perform an experiment with end-users to evaluate the RPHs through the privacy-solution produced by them. The criteria used to evaluate the design principles and the privacy solution are described in Section 5.

4. Improve and re-evaluate the approach: this step mainly focuses on identifying limitations or areas of improvement and refining the approach accordingly.

3.1 The process for constructing the approach

The design principles represent the cornerstone of our approach, and, as mentioned earlier, the first three steps of the approach design were dedicated to developing the design principles. However, to develop design principles, it is essential to understand their nature and purpose. Based on existing literature, Cronholm and Göbel [5] define design principles as clear, prescriptive statements that vary in their level of abstraction depending on the context. Their primary function is to encapsulate and communicate knowledge in a way that allows them to be reused in multiple instances that are under similar conditions [24]. Möller et al. [24] synthesize the definition of design principles as *"fundamental propositions that aid designers in achieving successful transfer of requirements to design"*.

With a clear definition of design principles, the next step in the artifact development process is to *elicit Meta-Requirements (MRs)* [24]. To do so, we recall the RPHs definition: RPHs are decision support mechanisms grounded on ethical principles such as *Respect*, *Beneficence (and Non-maleficence [17])*, *Justice*, *Integrity*, and *Social Responsibility* [31]. These ethical principles aim to empower user autonomy and enable informed decision-making.

Based on these foundations, we derive six meta-requirements that integrate ethical values while accounting for regulatory requirements, contextual adaptability, and accessibility. The complete list of MRs and their description are presented in Table 4.

Table 4. Meta-requirements for the design principles.

MR1. Integrity: information presented through RPHs must be accurate, truthful, consistent, and free from incomplete representations. [17, 1]
MR2. Non-manipulation: RPHs must not exploit cognitive biases or limitations to steer users toward privacy-invasive decisions.[1]
MR3. Beneficence and Non-maleficence: RPHs should maximize user benefits while minimizing privacy risks and potential harms, ensuring ethical and responsible guidance. [31, 17]
MR4. Autonomy and control: RPHs should empower users' autonomy and freedom of choice by offering genuine, meaningful, and informed options without implicit coercion. [31, 1, 17]
MR5. Context-aware and accessible: RPHs should, when possible, adapt to different contexts by considering risk levels, situational factors, and user diversity.
MR6. Regulatory compliant: RPHs should align with legal standards (e.g., GDPR).

The next step is to *formulate the design principles*. The formulation of RPHs design principles was grounded in the relevant literature identified in the *Identify knowledge base* step. Specifically, we examined existing deceptive patterns and related heuristics (see Table 2) and reviewed research on deceptive strategies (e.g., [21, 3, 12, 1]) to better understand how user behavior is manipulated or exploited, concerning their privacy choices. Additionally, these design principles have been formulated in light of the already mentioned meta-requirements.

We pay special attention to Ahuja and Kumar's [1] work, which identifies 25 dark strategies and then seven broad ethical concerns—compulsion, inadequate information, biased evaluation, insufficient deliberation, lack of control, pressure to conform, and restricted options—raised by these deceptive patterns. Their study further anchors these concerns into four theoretical conceptualizations of autonomy: agency, freedom of choice, control, and independence. Building on these insights, we formulated the design principles aimed at mitigating deceptive and unethical strategies while ensuring compliance with the established meta-requirements. We also derived acceptance criteria (AC) for each principle, to aid the designer during the RPHs evaluation.

AC serves as a checklist to help designers verify whether the design produced by RPHs satisfies the principles. They were formulated as *Yes* or *No* questions to facilitate this assessment. They were designed to be applied flexibly. Their practical applicability will be more comprehensible in the following section (Section 4), where we illustrate their use through an example. The principles and their associated acceptance criteria are organized in Table 5.

Table 5. Design principles and acceptance criteria for responsible privacy heuristics.

DP1.	**Neutral**: A RPH should present information about privacy choices in a neutral and balanced manner, avoiding framing that could lead to biased or skewed decisions.
DP1AC1.	Are all choices presented with equal prominence? (i.e., Are they displayed in a way to allow users to perceive them as equally relevant?)
DP2.	**Honesty and clarity**: A RPH should ensure that all information presented to users is truthful, clear, and easy to understand.
DP2AC1.	Is the privacy-related choice or information presented accurately and comprehensibly to users of different levels of expertise?
DP3.	**Navigable and actionable privacy**: A RPH should help users easily identify, understand, and act upon privacy-related information.
DP3AC1.	Is the path to privacy-related mechanisms easy to navigate to?
DP3AC2.	Are actionable privacy mechanisms (e.g., privacy settings) easily recognizable and intuitive to use?

DP4.	**Pressure-free**: A RPH should not impose time constraints, emotional manipulation, or other coercive tactics that pressure users into making privacy decisions.
DP4AC1.	Are users free to make privacy decisions without being subject to: time constraints; exclusive time-limited offers or other alleged financial gains in exchange for PI; coercive tactics exploiting emotional and social factors (e.g., guilt shaming, fear of missing out, and bandwagon effect)?
DP5.	**Benefit-risk balance**: A RPH should highlight user benefits and proactively minimize potential privacy risks.
DP5AC1.	Are the benefits and potential privacy risks associated with a given action clearly communicated?
DP6.	**Consequences awareness**: A RPH should provide feedback related to privacy choices, avoiding obscuring consequences, which could affect the users' decision-making.
DP6AC1.	Are the consequences of privacy choices clearly communicated to the user during or after a decision-making process, through real-time feedback or confirmation?
DP6AC2.	Is information about the implications of users' privacy choices easy to find?
DP7.	**Empowering**: A RPH should support users to select privacy choices that align with their privacy requirements.
DP7AC1.	Does the privacy solution provide intuitive and customizable privacy options that enable users to control, correct, and retract their privacy choices in a way that aligns with their preferences and requirements?
DP8.	**Context-aware**: A RPH should help users assess privacy decisions in context, considering factors such as data sensitivity, purpose of collection and use, recipient identity, and potential risks.
DP8AC1.	Does the privacy solution provide users with context-specific information that allows them to assess the sensitivity and risks implied by the type of data being requested (e.g., health, financial, or personal data), the purpose for its use, and the identity of the recipients?
DP9.	**Situation-aware**: A RPH should adapt to different situations to provide relevant, meaningful, and actionable guidance.
DP9AC1.	Is privacy guidance provided based on the user's current situation (e.g., location, device type, or task being performed) and interaction context (e.g., signing up for a service vs. sharing a photo)?
DP10.	**Accessible and inclusive**: A RPH should ensure that users, regardless of their abilities or technical expertise, can understand and act upon privacy-related information.

DP10AC1.	Is privacy-related information provided in multiple formats (e.g., simple language, assistive technologies, alternative formats) to accommodate users' varying preferences, expertise, sensory needs, and disabilities?
DP11.	**Regulation Compliant:** A RPH must not encourage or lead to violating privacy legislation (e.g., purpose limitation, data minimization).
DP11AC1.	Are privacy-related choices and information compliant with the relevant privacy legislation (e.g., GDPR in Europe)?

As mentioned previously, the design principles were formulated in compliance with the meta-requirements, which cover ethical values and regulatory requirements. Table 6, presents a matrix linking each design principle to the associated meta-requirements. Next, we comment on how each principle complies with these meta-requirements.

Table 6. Meta-requirements satisfied by the design principles.

	MR1	**MR2**	**MR3**	**MR4**	**MR5**	**MR6**
DP1	X	X	X			
DP2	X	X	X			X
DP3		X	X	X		
DP4		X	X	X		
DP5		X	X			
DP6	X	X	X			
DP7		X	X	X		
DP8		X	X	X	X	X
DP9		X	X	X	X	
DP10		X	X		X	X
DP11	X	X	X			X

The following design principles are compliant with **MR1. Integrity**: **DP1. Neutral** by requiring RPHs to be balanced and free of framing; **DP2. Honesty and clarity** by requiring RPHs to be truthful and clear; **DP6. Consequence awareness** by requiring RPHs to provide complete and accurate information regarding the user's choices; and **DP11. Regulation compliant** by requiring RPHs to comply with regulations, such as Europe's GDPR, which include providing information in a transparent manner [8].

With respect to **MR2. Non-manipulation** and **MR3. Beneficence and Non-maleficence**, all principles are compliant. The principles encourage RPHs to present privacy-related information in an unbiased, accurate, and accessible manner (e.g., **DP1. Neutral, DP2. Honesty and clarity**, and **DP10. Accessible and inclusive**). They present actionable mechanisms in a way that makes it intuitive for DSs to learn and act upon them (e.g., **DP3. Navigable and actionable**), and comprehend their repercussions (e.g., **DP6. Consequence-awareness**). They also guide RPHs to enable decisions free

of constraints (e.g., **DP4. Pressure-free**), and empower the DS to make these decisions according to their preferences (e.g., **DP7. Empowering**). In addition, the principles require RPHs provide the DS with information tailored to the situation, risk level, and regulatory expectations (e.g., **DP8. Context-aware, DP9.Situation-aware** and **DP11. Regulation compliant**). In general, these principles work to reduce cognitive load, prevent manipulative practices, and support informed, autonomous decision-making, thus helping mitigate potential privacy risks.

For **MR4. Autonomy and control**, the following principles comply: **DP3. Navigable and actionable privacy** by requiring that RPHs provide mechanisms that are easy to recognize and to use; **DP4. Pressure-free** by requiring RPHs to allow deliberate and voluntary choices, free of constraints; **DP7. Empowering** by requiring RPHs to support choices that align with users preferences; **DP8. Context-aware** by requiring RPHs to help users assess privacy choices, while considering the context in which they are being made and PI that is being required; and **DP9. Situation-aware** by encouraging RPHs to provide meaningful guidance allowing users to adapt and make informed decisions.

The principles that comply with **MR5. Context-aware and accessible** are: **DP8. Context-aware** by requiring RPHs to provide information concerning data sensitivity, purpose of collection and use, and potential risks; **DP9. Situation-aware** by requiring RPHs to be adaptable to different situations, providing users with meaningful guidance; and **DP10. Accessible and inclusive** by requiring RPHs to enable accessible and inclusive design.

Finally, the principles compliant with **MR6. Regulatory compliant** are the following: **DP2. Honesty and clarity** by requiring RPHs to offer truthful and accurate information; **DP8. Context-aware** by requiring RPHs to help users understand factors such as purpose of collection and use, which relates to the transparency required by regulations (e.g., GDPR [7]); **DP10. Accessible and inclusive** by requiring RPHs to provide information in an intelligible and easily accessible manner, which relates to the *"Rights of the Subject"* outlined in Europe's GDPR [8]; and **DP11. Regulation compliant** which requires RPHs to stay compliant with regulations.

3.2 The Methodological Process

In this subsection, we outline the methodology to be followed for designing RPHs. The process is an improved version of [30][1], illustrated in Figure 3, consists of four key steps:

1. **Identify core privacy heuristics for usability:** takes the privacy solution (system) as input and derives PHs that enhance its usability. The goal is to ensure that privacy-related interactions are intuitive, clear, and user-friendly, making it easier for users to understand and manage their privacy settings effectively. We apply

[1]This paper is an earlier version of the overall approach.

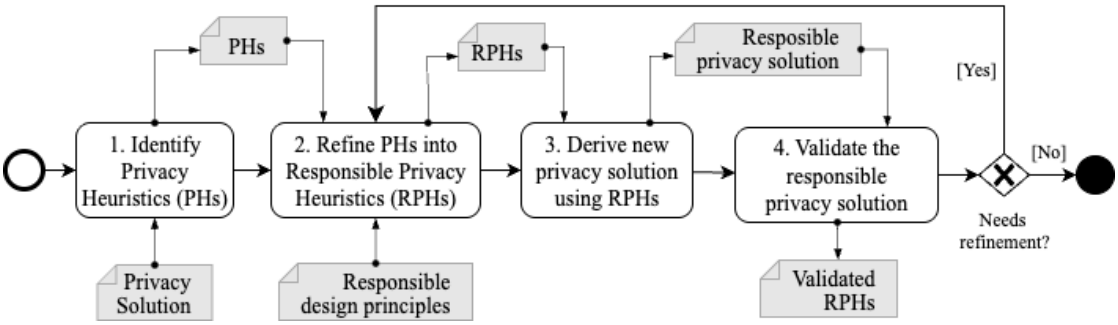


Figure 3. The methodological process to be followed during the overall RPHs design.

Gharib's 10 Usable Privacy Heuristics (UPHs[2]) [10], which are listed in Table 7, to guide the design process. Note that we kept the identifier as PH to stay consistent with how we've been referring to privacy heuristics throughout this work.

2. **Refine PHs into RPHs:** this step takes the identified PHs and the responsible design principles as input. The designer then selects one or more principles that best enhance the PHs to develop them into RPHs. Throughout this process, the designer may consult the corresponding acceptance criteria to refine and strengthen the resulting RPHs.

3. **Derive a new privacy solution based on RPHs:** this step uses the selected RPHs and their corresponding acceptance criteria to guide the design of a revised privacy solution. During the design process, the designer evaluates the design using the acceptance criteria. To facilitate this evaluation, the acceptance criteria are phrased as yes-or-no questions. If the privacy solution meets the criteria with positive responses, the principles used to derive the RPH is considered satisfied.

4. **Validate the responsible privacy solution:** evaluates whether the RPHs achieve the purpose of their development, which can be carried out using one or a combination of commonly used methods, such as end-user testing, expert reviews, or other assessment techniques.

[2]In the context of this work we treat privacy heuristics (PHs) and usable privacy heuristics (UPHs) as synonyms since the main objective of PHs is making privacy solutions usable.

Table 7. Usable privacy heuristics and their acceptance criteria [10].

PH1	**Visibility**: A DS should be informed about their privacy choices.
PH1AC	Is there any feedback for every privacy-related action?
PH2	**Revocability**: A DS should be allowed to revoke any privacy actions.
PH2AC	Can DSs easily reverse their privacy actions?
PH3	**Clarity**: A DS should be informed about the consequences of any privacy actions.
PH3AC	Does the system warn DSs if they are about to make a potentially privacy error?
PH4	**Expressiveness**: A DS should be guided on privacy while still being able to have freedom of expression.
PH4AC	Is there a clear understanding of the system's privacy options?
PH5	**Learnability**: A DS should be ensured that privacy actions are easy to learn and remember.
PH5AC	Are privacy operations easy to learn and use?
PH6	**Minimalist design**: A DS should be offered relevant information relating to their privacy actions.
PH6AC	Is only the privacy information essential to decision-making displayed to the user?
PH7	**Errors**: A DS should be provided with detailed privacy error messages that they can understand and act upon
PH7AC	Do error messages suggest the cause of the privacy problem, and how it can be corrected?
PH8	**Satisfaction**: A DS should be provided with a good experience when making a privacy decision and ensured that they are in control
PH8AC	Do privacy-related prompts imply that the user is in control?
PH9	**User suitability**: DSs should be provided with options considering their diverse levels of skill and experience in security.
PH9AC	If the system supports both novice and expert DSs, are multiple levels of privacy error messages available?
PH10	**User assistance**: A DS should have access to apparent privacy help.
PH10AC	Is there a visible privacy help?

4 Applying the approach to examples from the online social network domain

In this section, we demonstrate the application of the approach by redesigning the privacy settings interface of a social media platform. Privacy settings serve as one of the primary tools through which DSs can manage and control their personal information, enabling real-time decisions about how their data is shared, accessed, and processed.

To better understand the types of actions and information that online social network (OSN) privacy settings typically provide, it is essential to first consider their core functionalities. We reviewed widely used OSNs—such as Facebook and Instagram (via Meta's Privacy Center), LinkedIn, and Reddit—to identify recurring features and patterns in their functionalities and privacy controls. Based on this analysis, we draw the scope of the OSN example and outline their privacy settings.

In general, OSNs enable users to create and personalize profiles, which typically includes details such as name, profile picture, description, birthday, and email address. Users can share content through posts that combine visual and textual elements (e.g., photos with descriptions and tags), interact with content (their own, others', or advertisements) via reactions, comments, and sharing, and engage in direct communication through messaging features. Table 8 shows the privacy settings that correspond with the functionalities of the example.

Table 8. Key privacy settings of OSN based on the scope defined.

Profile Visibility: controls the visibility of profile details (e.g., profile picture, description, birthday, email, friends list, profiles the user follows, tagged content), and activities (e.g., user's posts, and whether their comments and reactions to other's posts show up on their friends feeds).
Account & Security: controls over account details (e.g., change email used for account recovery, change password, enable/disable two-factor authentication); account deletion and temporary deactivation, and active sessions management;
Interaction Preferences: controls how others can interact with DS's profile (e.g., who can tag the DS or comment on their posts, and who can message them), and activity (e.g., who can interact with users' posts, and blocking profiles);
Ad Preferences: control and information over ads customization and manage experience (e.g., what information can be accessed and processed, learn who uses and what data they use on advertisement customization);
Permissions and Policies: control over data access and processing (e.g., service provider and third-party permissions), and policies (e.g., privacy policy, cookie policy, and terms and conditions);

We selected the first group of settings from Table 8, **Profile Visibility**, to demonstrate the applicability of our approach. To provide a concise example, we chose to design five interfaces concerning **Profile Visibility**, and treat each one of them as a distinct privacy solution. Table 9 presents these privacy solutions along with brief descriptions.

Table 9. Privacy solutions used in the demonstration of the approach.

Profile visibility default interface: This is the main screen of the Profile Visibility settings. From this interface, users can navigate to individual visibility settings.
Who can see your profile description setting: This interface allows users to control who can view their profile description, which includes information such as workplace, education, and locations.
Who can see your tagged content setting: This interface enables users to manage who can view posts they are tagged in, when accessed through their profile. Note that these posts may still be visible via other sources (e.g., another user's profile).
Who can see your posts setting: This interface allows users to control which groups of people can view the content they post.
Choose if your reactions or comments show up on your friends' feed setting: This interface provides two privacy controls, one for managing the visibility of reactions and another for comments. These controls only affect whether such actions appear on friends' feeds; friends may still see them by visiting the original post.

In the next subsection, we go through the methodological process for each of the privacy solutions presented above. We highlight that since these settings belong to the same group, many of the design principles applied may be similar or even identical across privacy solutions. In such cases, designers may choose to skip the refinement step (**Step 2**) if the RPHs derived for one solution also adequately address the needs of another. They may also reuse a RPH from one solution and refine it further to meet the need of another. These alternatives will be shown in the next subsection.

4.1 Profile visibility default interface

In **Step 1**, we take the default interface of the *Profile Visibility* settings as our first privacy solution, illustrated in Figure 4. Through which users can access settings to change the visibility of profile details and activities. Then, we identify the core privacy heuristics that enhance usability.

To improve usability, we applied a minimalist and consistent layout that provides data subjects with essential information for their privacy actions following **PH6. Minimalist design:** A DS should be offered relevant information relating to their privacy actions. At the top of the interface (see Figure 4), the DS is met with a brief description of what

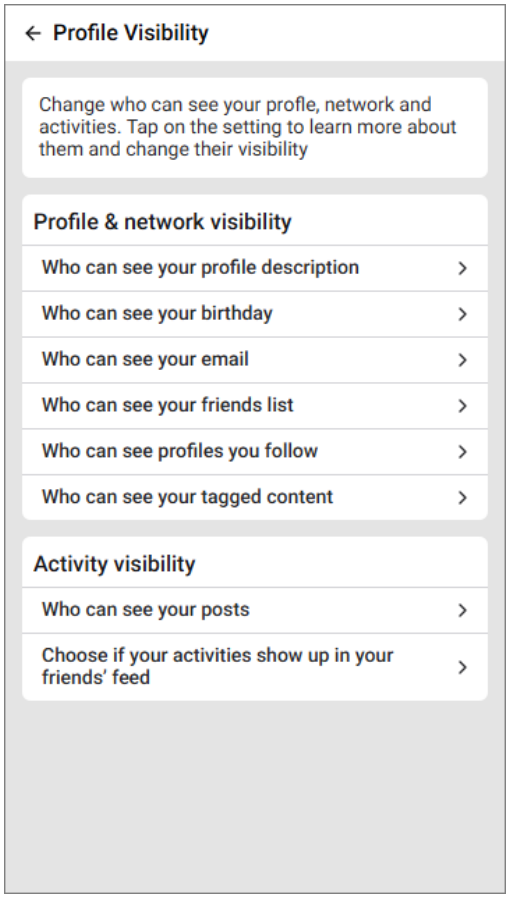


Figure 4. Profile visibility settings following PH.

these settings enable. They are also informed that tapping on the setting will lead them to more information about it, inspired by **PH4. Expressiveness:** A DS should be guided on privacy while still being able to have freedom of expression; and **PH1. Visibility:** A DS should be informed about their privacy choices.

After the description, the privacy settings are split into two sections, with descriptive names that avoid technical terms. This design choice takes into account that DSs have different levels of digital literacy and familiarity with privacy settings, as suggested by **PH9. User suitability:** DSs should be provided with options considering their diverse levels of skill and experience in security.

In **Step 2**, we refine the PHs into RPHs by analyzing each PH individually and applying the design principles deemed appropriate. This refinement process is documented in Table 10. For each PH, we first present its original form, followed by its refined version, where the changes introduced by applying each principle are highlighted in **bold** (e.g.,

RPH1.1[3]).

Table 10. Step 2 of the methodological process for profile visibility default interface.

PH1. Visibility	"A DS should be informed about their privacy choices."
RPH1.1 refined by DP9. Situation-aware	"A DS should be **provided with meaningful and actionable guidance, adapted to the interaction context, when** informed about their privacy choices."
RPH1.2 refined by DP8. Context-aware	"A DS should be provided with meaningful, actionable guidance, adapted to the interaction context **and considerate of the data sensitivity**, when informed about their privacy choices, **allowing them to recognize potential privacy risks.**"
PH4. Expressiveness	"A DS should be guided on privacy while still being able to have freedom of expression."
RPH4.1 refined by DP1. Neutral	"A DS should be **provided with unbiased and balanced information about** privacy while still being able to have freedom of expression."
PH6. Minimalist design	"A DS should be offered relevant information relating to their privacy actions."
RPH6.1 refined by DP3. Navigable and actionable privacy	"A DS should be offered relevant**, easy to learn information** relating to their privacy actions **, making sure it is easily recognizable and usable.**"
PH9. User suitability	"DSs should be provided with options considering their diverse levels of skill and experience in security."
RPH9.1 refined by DP10. Accessible and inclusive	"DSs should be provided with **inclusive** options considering their diverse levels of skill **and accessibility needs.**"

Now that PHs have been refined into RPHs, we move to **Step 3** of the process by revisiting the initial solution to update the interface accordingly, shown in Figure 5. We analyze the design, from top to bottom, addressing the changes introduced and associated RPHs they reflect.

The description has been revised into a question with a slightly bolder text to capture the DSs' attention. Although the content remains the same as the previous version, it has been broken down into two bullet points to emphasize the types of actions these settings enable. This change is guided by **RPH6.1**, which encourages minimal design that is easy

[3]To maintain traceability, each RPH retains the number of its originating PH. The number after the dot indicates the version, incremented each time a different design principle is applied. For example, **RPH1.1** and **RPH1.2** are both refinements of **PH1. Visibility**

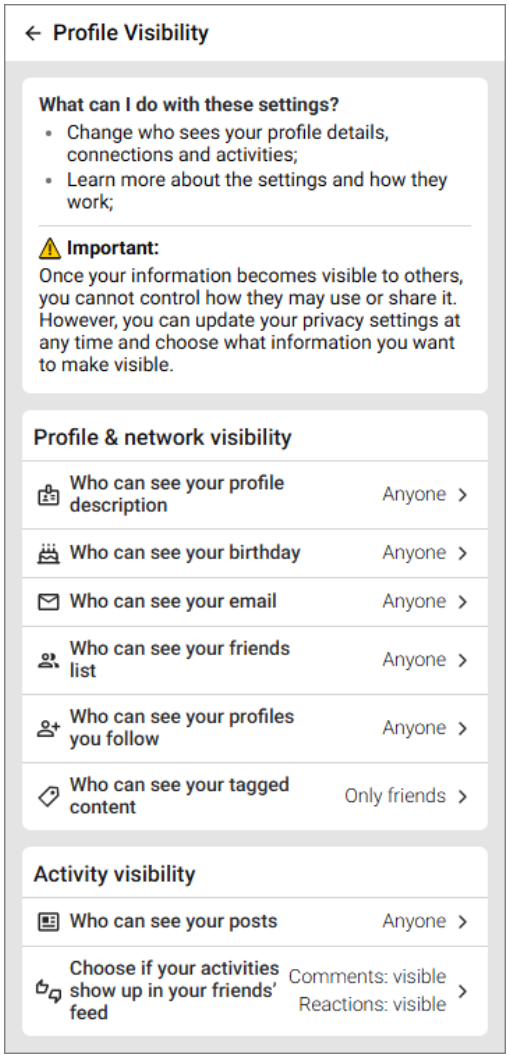


Figure 5. Profile visibility settings following RPHs.

to navigate, recognize, and use. In addition, a warning has been added to inform DSs about potential privacy risks associated with the information they choose to make visible. This aligns with **RPH1.2**, which promotes visibility of information through meaningful guidance, helping DSs reflect on possible risks that might come with making personal information visible.

After the description, the two-section structure remains the same as in the previous version, but the settings descriptions have been enhanced with visual elements (e.g., icons), and the current setting status is now visible without requiring DSs to open each setting. The addition of icons was guided by **RPH6.1**, which enhances navigation

through recognizable elements, and **RPH9.1**, which reinforces the need for inclusive and accessible information. When properly used, visual elements can support people with cognitive disabilities and, more broadly, assist DSs in locating and recognizing the purpose of each control.

By displaying the setting statuses directly on the *Profile Visibility* default interface we enable DSs to quickly go through current privacy choices and assess whether any changes are needed. This change was done to set expectations about the type of privacy options provided, and improve user engagement with the settings. Both which are grounded in **RPH9.1**, **RPH6.1** and **RPH1.2**. Additionally, although the design contains more information, we try to keep it balanced in order not to overwhelm the DS, while being careful not to favor specific privacy actions, as suggested by **RPH4.1**, which supports the DS through balanced unbiased design.

The final part of this step involves evaluating the design against the acceptance criteria of the applied principles. To facilitate this assessment, we revisit the relevant acceptance criteria (previously introduced in Table 5) and organize them in a new table with corresponding evaluations below each of them. See Table 11.

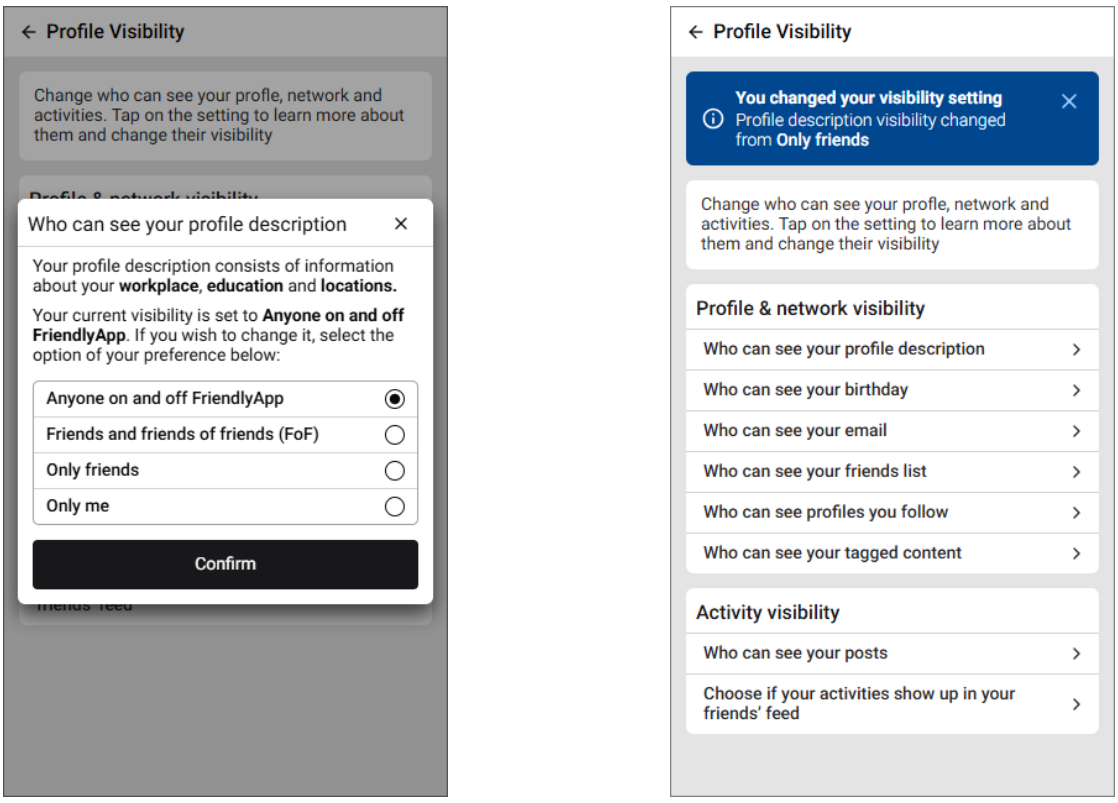
Table 11. Profile visibility default interface evaluation with AC.

DP1AC1. Are all choices presented with equal prominence? (i.e., Are they displayed in a way to allow users to perceive them as equally relevant?)
Yes. All links to the privacy settings follow the pattern: icon, name, and current status.
DP3AC1. Is the path to privacy-related mechanisms easy to navigate to?
Yes. The privacy settings are one click away from this interface, and their links were updated with recognizable visual elements (e.g., icons) and current setting status.
DP3AC2. Are actionable privacy mechanisms (e.g., privacy settings) easily recognizable and intuitive to use?
Yes. Icons were chosen to resemble the functionality their are associated (e.g., a birthday cake icon for birthday visibility). The inclusion of setting statuses upfront helps set user expectations about the controls and whether changes are needed.
DP8AC1. Does the privacy solution provide users with context-specific information that allows them to assess the sensitivity and risks implied by the type of data being requested, the purpose for its use, and the identity of the recipients?
Yes. Interpreting *"data being requested"* as *"data being made visible"* the new design explicitly warns users that others may use their information in unexpected ways. The privacy risks were not explicitly mentioned to avoid pressuring the DS.
DP9AC1. Is privacy guidance provided based on the user's current situation (e.g., location, device type, or task being performed) and interaction context (e.g., signing up for a service vs. sharing a photo)?

Yes. The description at the top and the warning serve to guide the DS on how to use these controls and what to keep in mind, given the interaction context (i.e., control of profile information visibility).
DP10AC1. Is privacy-related information provided in multiple formats (e.g., simple language, assistive technologies, alternative formats) to accommodate users' varying preferences, expertise, sensory needs, and disabilities?
Yes. The enhanced design maintains simple language, devoid of technical terms; utilize different font weights to create contrast, and visual elements to make the design more engaging and sections of information more identifiable.

4.2 Who can see your profile description

Starting from **Step 1**, the second interface we design is pertinent to a specific setting, named *"Who can see your profile description"*, shown in Figure 6a.



(a) (b)

Figure 6. *Who can see your profile description* (a) setting and (b) feedback, PH version.

This design follows several PHs introduced in the previous example. It supports **PH1.**

Visibility by clearly informing DSs of their privacy choices; adheres to **PH6. Minimalist Design** by maintaining a simple layout that presents only relevant information about the privacy choices; and promotes **PH9. User Suitability** through the use of simple, accessible language.

Because this interface facilitates a privacy-related decision, it also addresses the ease of learning and use, in line with **PH5. Learnability:** A DS should be ensured that privacy actions are easy to learn and remember. Furthermore, the inclusion of a feedback dialog (see Figure 6b) serves to clarify the outcomes of DSs' decisions, thereby aligning with **PH3. Clarity:** A DS should be made aware of the implications of their privacy actions.

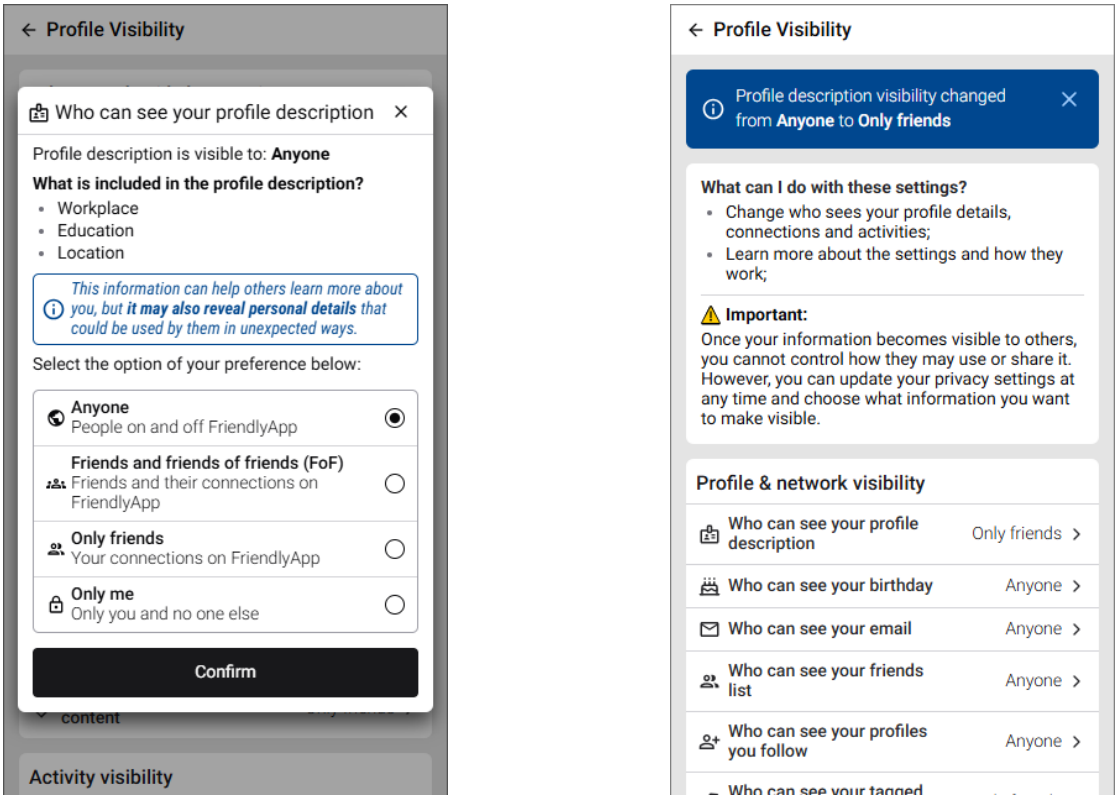
In **Step 2**, we refine the relevant PHs into RPHs. We identify that **RPH1.2** and **RPH9.1** appropriately address the needs of this design, making it unnecessary to refine **PH1. Visibility PH9. User Suitability** again. Similarly, we observe that **RPH6.1** provides a broader yet compatible interpretation of **PH5. Learnability**, and thus, there is no need for a refinement of **PH5. Learnability**. However, given that the current interface introduces specific privacy actions, we see the need to further refine **RPH6.1**. See Table 12:

Table 12. Step 2 of the methodological process for *Who can see your profile description*.

PH3. Clarity	"A DS should be informed about the consequences of any privacy actions."
RPH3.1 refined by DP6. Consequences awareness	"A DS should be informed about the consequences of any privacy actions **before and after a decision, through meaningful real-time feedback.**"
RPH3.2 refined by DP4. Pressure-free	"A DS should be informed about the consequences of any privacy actions before and after a decision, through meaningful real-time feedback**, in a manner that does not pressure or coerce users**."
PH6. Minimalist design	"A DS should be offered relevant information relating to their privacy actions."
RPH6.1 refined by DP3. Navigable and actionable privacy	"A DS should be offered relevant**, easy to learn information** relating to their privacy actions **, making sure it is easily recognizable and usable.**"
RPH6.2 refined by DP1. Neutral	"A DS should be offered relevant, easy to learn information relating to their privacy actions, **presented in a neutral and balanced way,** making sure it is easily recognizable and usable**, as well as free from biased framing**."

We emphasize that there is flexibility in the use of these principles. If the designer sees fit, they may refine these PHs using different principles or even build on top of the previous RPH. Alternatively, a PH does not need to be refined again if it is already satisfied by an existing RPH.

In **Step 3**, we revisit the design applying the RPHs, resulting in the designs illustrated in Figures 7a and 7b.



(a) (b)

Figure 7. *Who can see your profile description* (a) setting and (b) feedback, RPH version.

The updated interface (Figure 7a) maintains the current setting status at the top of the modal, with minor rewording to enhance clarity. It also rearranges the explanation into a question and bullet points to emphasize what is included in the profile description. In the following, an information box (outlined in blue) highlights that this type of information can reveal personal details and that the application cannot anticipate how others might use this information.

These changes along with the additional information box was guided by **RPH6.2**, as it ensures relevant information is presented in a unbiased and comprehensible manner; also **RPH1.2** and **RPH3.2**, by providing meaningful and contextually relevant guidance

that allows the user to reflect on the possible consequences, without feeling pressured to pick a specific option.

Privacy options were improved by including icons and brief descriptions to increase clarity and make the interface more intuitive. It also presents these options with equal prominence, avoiding biased framing. This enhancement aligns with **RPH6.2**, which emphasizes that information on privacy options should be balanced, easily recognizable and intelligible. It is also informed by **RPH9.1**, which advocates for an inclusive and accessible design. The latter is reflected in the use of plain language, typographic contrast (e.g., bold and light text), and color cues to help users identify key elements such as the information box.

After confirming their choice, DSs are presented with a feedback dialog (Figure 7b), which has been slightly modified for improved clarity. The feedback message has been rephrased to clearly specify which setting was changed (e.g. profile description), what was the previous visibility status (e.g., Anyone), and what is the current (e.g., Only friends). This modification is guided by **RPH3.2**, which emphasizes the importance of informing DSs about the consequences of their privacy actions through meaningful, real-time feedback, without applying any pressure or unnecessary constraints.

Finally, we evaluate the new design against the acceptance criteria of the applied principles in the same way as in the first example. Table 13.

Table 13. *Who can see your profile description* setting evaluation with AC.

DP1AC1. Are all choices presented with equal prominence? (i.e., Are they displayed in a way to allow users to perceive them as equally relevant?)
Yes. Although the information box signals possible risks the information provided is unbiased, and the privacy options follow the same design.
DP3AC1. Is the path to privacy-related mechanisms easy to navigate to?
Yes. The interface has been enhanced to make the information easy to learn, and privacy options more intuitive to use.
DP3AC2. Are actionable privacy mechanisms (e.g., privacy settings) easily recognizable and intuitive to use?
Yes. The privacy options were kept in the radio button format, but their design has been enhanced with icons and descriptions to improve recognition.
DP4AC1. Are users free to make privacy decisions without being subject to: time constraints; exclusive time-limited offers or other alleged financial gains in exchange for PI; coercive tactics exploiting emotional and social factors (e.g., guilt shaming, fear of missing out, and bandwagon effect)?
Yes. Although the interface has an information box aiming to make the user reflect on possible risks, we do not pressure them to choose or avoid a specific privacy option.

DP6AC1. Are the consequences of privacy choices clearly communicated to the user during or after a decision-making process, through real-time feedback or confirmation?
Yes. As soon as the DS confirm their changes, they are met with a feedback dialog reminding them of the change. It shows both the previous status and the current one.
DP6AC2. Is information about the implications of users' privacy choices easy to find?
Yes. They are informed about what others, excluded from their selection, can view, and receive feedback whenever they make a change in the setting.
DP8AC1. Does the privacy solution provide users with context-specific information that allows them to assess the sensitivity and risks implied by the type of data being requested, the purpose for its use, and the identity of the recipients?
Yes. The new design highlights that profile description contains personal details that can be used by others in ways the app cannot anticipate, to prompt DSs to reflect more on their choices.
DP9AC1. Is privacy guidance provided based on the user's current situation (e.g., location, device type, or task being performed) and interaction context (e.g., signing up for a service vs. sharing a photo)?
Yes. The design highlights that other might use their information in unexpected ways, so the user can reflect on the risks before deciding on an option.
DP10AC1. Is privacy-related information provided in multiple formats (e.g., simple language, assistive technologies, alternative formats) to accommodate users' varying preferences, expertise, sensory needs, and disabilities?
Yes. Different font weights and elements (e.g., information box) create contrast, to make the design more engaging and sections of information more identifiable.

4.2.1 Who can see your tagged content

This setting design is very similar to the previous one (i.e., *Who can see your profile description*), hence we consider all the PHs used in it. The only addition is **PH4. Expressiveness**: A DS should be guided on privacy while still being able to have freedom of expression. This PH guided the inclusion of a note to inform DSs about the limitations of this setting and to highlight additional ways to control the visibility of tagged content. The resulting design is illustrated in Figure 8. We emphasize that the feedback dialog follows the same pattern as the previous setting, so we choose to omit it.

For **Step 2** we reuse most RPHs used in the previous setting, namely: **RPH3.2**, **RPH6.2**, **RPH9.1**; and **RPH1.1** from *Profile visibility default interface*. The only PH left for refinement is **PH4. Expressiveness**. However, this PH is already addressed by **RPH1.1** and **RPH3.2**, which informs that the design must provide meaningful guidance

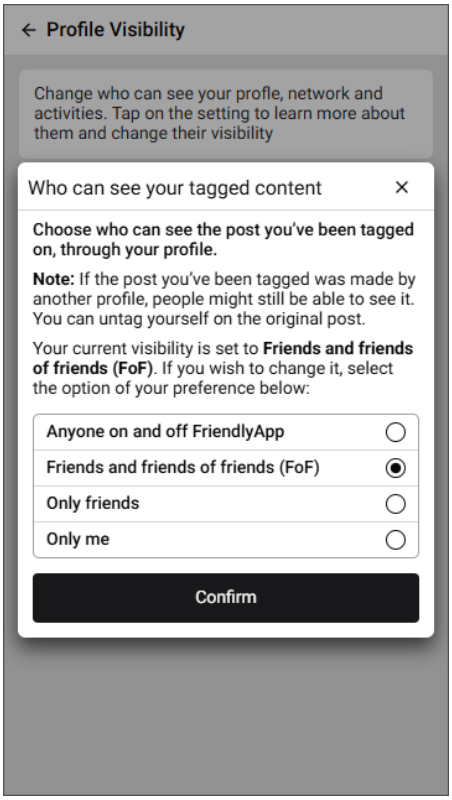


Figure 8. *Who can see your tagged content*, PH version.

on information about privacy choices and their consequences, while making sure the user is free to make their choice.

In **Step 3** we revisit the design and enhance it using the RPHs (see Figure 9). To keep a consistent layout across the settings, we follow the same pattern for the description, before the privacy options: start off with the current status of the setting, then a brief explanation about the setting or functionality, and finally any additional information about its limitations and/or further instructions.

We apply this pattern consistently across other settings within the group (i.e., **Profile Visibility**), as we believe that this consistency will enhance DSs ability to navigate and use these settings effectively. This decision is grounded in **RPH6.2**, which advocates for a minimalist design that avoids biased framing and promotes privacy actions that are easy to learn, recognize, and use. Furthermore, it draws on **RPH9.1**, which calls for accessible and inclusive design, in this case, by establishing consistent expectations for how information is presented throughout the interface.

To support DSs who may be unfamiliar with the concept of tagged content, the interface includes a brief explanation of the functionality. Below, an information box

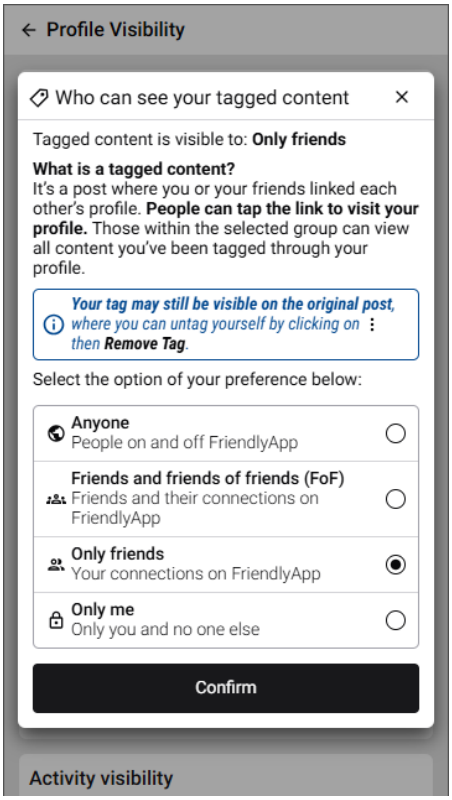


Figure 9. *Who can see your tagged content*, RPH version.

(outlined in blue) clarifies that this setting only controls the visibility of tagged content on the DS's profile, while the tag itself may still be active and visible in the original post. Moreover, the interface provides a short instructional note on how DSs can remove the tag themselves if desired.

This approach is guided by **RPH9.1** and **RPH6.2**, which emphasizes inclusive and accessible design that accounts for users' varying levels of familiarity with social media settings and features, and provides the DS with unbiased and comprehensible information that supports informed decision-making, respectively. Making limitations explicit and offering guidance on how to overcome them aligns with **RPH1.1**, which supports meaningful and actionable guidance that considers the DS current interaction context.

Finally, we evaluate the new design against the acceptance criteria of the applied principles, following the same approach as in the first example. Since this setting employs the same principles as the previous setting (i.e., Who can see your profile description) and both share similar design choices, we refer back to Table 13 for the ACs not explicitly discussed in Table 14.

Table 14. *Who can see your tagged content* setting evaluation with AC.

DP9AC1. Is privacy guidance provided based on the user's current situation (e.g., location, device type, or task being performed) and interaction context (e.g., signing up for a service vs. sharing a photo)?
Yes. After the explanation about what is tagged content, we specific mention that this setting controls who can view it when visiting the user's profile. Additionally, we enhance the instructions on how to untag oneself from a post by providing the steps to achieve it and making them more visible.
DP10AC1. Is privacy-related information provided in multiple formats (e.g., simple language, assistive technologies, alternative formats) to accommodate users' varying preferences, expertise, sensory needs, and disabilities?
Yes. In addition to enhancing the design with visual elements, an explanation of what a tagged content is was added for those not familiar with this feature.

4.2.2 Who can see your posts interface

Starting from **Step 1** we apply the same PHs from the previous setting: **PH1. Visibility**, **PH3. Clarity**, **PH4. Expressiveness**, **PH6. Minimalist Design**, and **PH9. User suitability**.

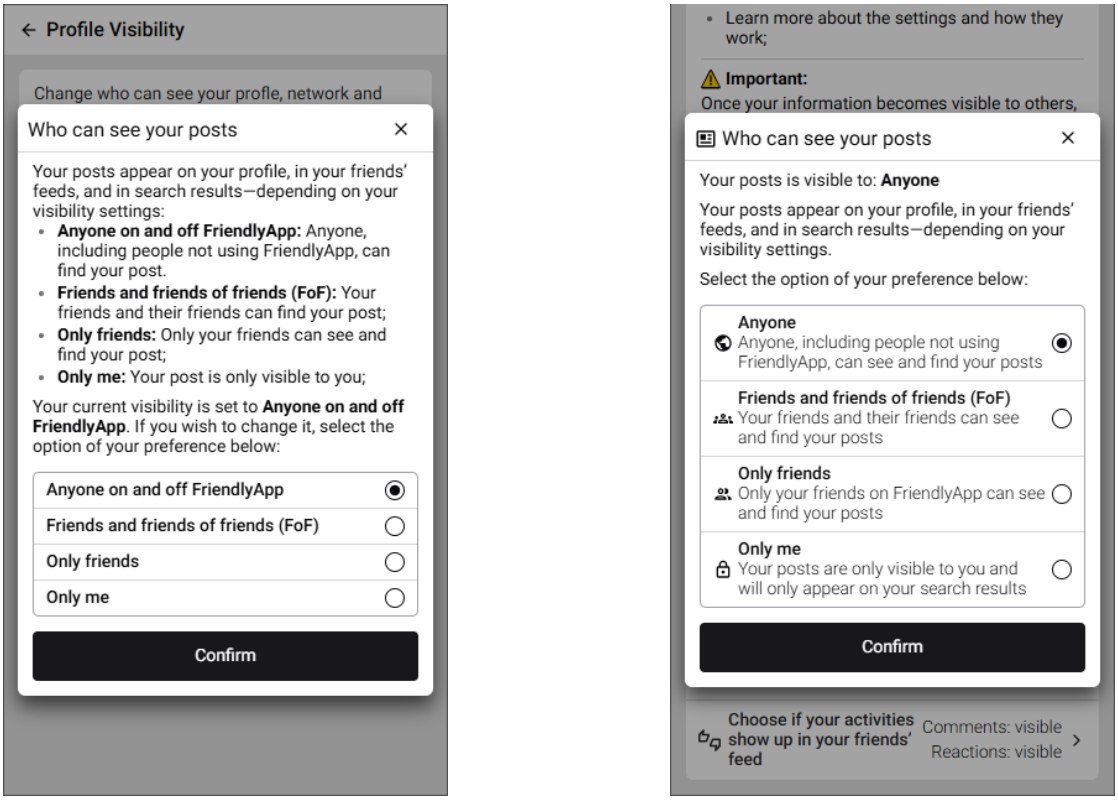
For **Step 2** we use some of the RPHs developed previously, namely: **RPH1.1**, **RPH3.2**, **RPH6.2** and **RPH9.1**. However, for **RPH1.1** we choose to refine it further with **DP2. Honesty and Clarity** to emphasize the need for a clear and correctly placed information. See Table 15.

Table 15. Step 2 of the methodological process for *Who can see your posts*.

PH1. Visibility	"A DS should be informed about their privacy choices."
RPH1.1 refined by DP9. Situation-aware	"A DS should be **provided with meaningful and actionable guidance, adapted to the interaction context, when** informed about their privacy choices."
RPH1.3 refined by DP2. Honesty and Clarity	"A DS should be provided with **clear,** meaningful and actionable guidance, adapted to the interaction context **and situated appropriately within it**, when informed about their privacy choices."

Both the initial and refined interfaces are presented side by side (see Figures 10a and 10b) to facilitate comparison. We briefly discuss the distinct elements of each version, highlighting the relevant PHs and RPHs applied. Design decisions that overlap with those of previous settings are not reiterated, as they have already been addressed.

In the PHs version (Figure 10a), the setting starts by explaining what post visibility entails, namely that posts may appear on the DS's profile, in their friends' feeds, and



(a) (b)

Figure 10. *Who can see your posts* (a) PH version and (b) RPH version.

in search results, depending on the selected visibility level. This general explanation is then expanded upon for each privacy option. This design reflects **PH1. Visibility**: A DS should be informed about their privacy choices; and **PH4. Expressiveness**: A DS should be guided on privacy while still being able to have freedom of expression.

In the updated design (Figure 10b), the explanation of visibility implications has been embedded directly within each privacy option. This allows users to immediately understand how each choice affects the visibility of their posts, without the need to shift focus between options and their consequences. This design decision is inspired by: **RPH1.3**, through the clear guidance embedded in the privacy option; **RPH6.2**, through enhanced description of privacy options to improve learnability and use, while keeping them unbiased and concise; and **RPH9.1** by making sure the instructions are described in a simple language.

The feedback interface for this setting follows the same structure as those presented in the previous two settings, with the only difference being the name of the setting. Therefore, we opt to omit its depiction and refrain from further discussion.

To conclude **Step 3**, we evaluate the design against the acceptance criteria of the

applied principles. This evaluation process results in conclusions that closely mirror those presented in the previous setting (e.g., Who can see your profile description), shown in Table 13. For that reason we omit ACs that were evaluated in a similar fashion and emphasize on the ones with different explanation. See Table 16.

Table 16. *Who can see your posts* setting evaluation with AC.

DP2AC1. Is the privacy-related choice or information presented accurately and comprehensibly to users of different levels of expertise?
Yes. We first provided a clear explanation about where the DS's posts can appear, then slightly expanded that explanation in the privacy option description. We avoided technical language or overwhelming the DS with information.
DP9AC1. Is privacy guidance provided based on the user's current situation (e.g., location, device type, or task being performed) and interaction context (e.g., signing up for a service vs. sharing a photo)?
Yes. Although the privacy options are the same as in the previous two settings, their descriptions have been enhanced to help users understand the specifics of this setting, without the need to switch back and forth between the information at the top and the related options.

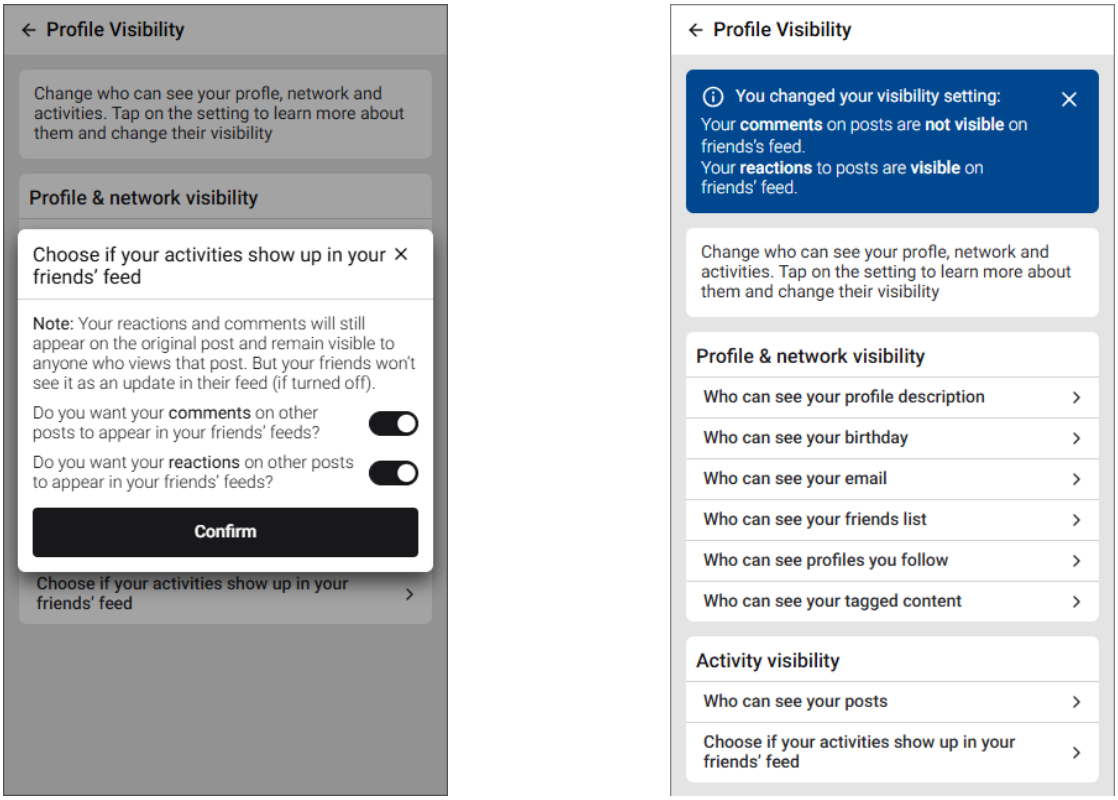
4.2.3 Choose if your reactions or comments show up on your friends' feed interface

We go through the first three steps of the process once more for the setting *Choose if your reactions or comments show up on your friends' feed*. Unlike the previous settings, this interface manages two functionalities, reactions and comments visibility, and does not follow the same set of visibility options as the previous settings (see Figure 11a). The feedback dialog layout is very similar to other settings, but informs two changes (see Figure 11b).

The interface is kept simple, featuring a notice that informs DSs that this setting does not alter the visibility of reactions or comments on the original post itself. Instead, it controls whether these activities will appear in their friends' feeds. This design decision is guided by **PH1. Visibility**: A DS should be informed about their privacy choices; and **PH4. Expressiveness**: A DS should be guided on privacy while still being able to have freedom of expression.

To maintain consistency with the other settings, we also apply the same PHs previously considered, namely: **PH3. Clarity**, **PH6. Minimal Design**, and **PH9. User Suitability**. We adopt the same RPHs as the *Who can see your tagged content* setting. As no additional refinements are necessary in this case, we proceed directly to **Step 3**, where we discuss the design changes using the selected RPHs.

Figure 12a presents the updated design, which differs significantly from the initial version. The layout adheres to the pattern shown in the previous settings: it begins with



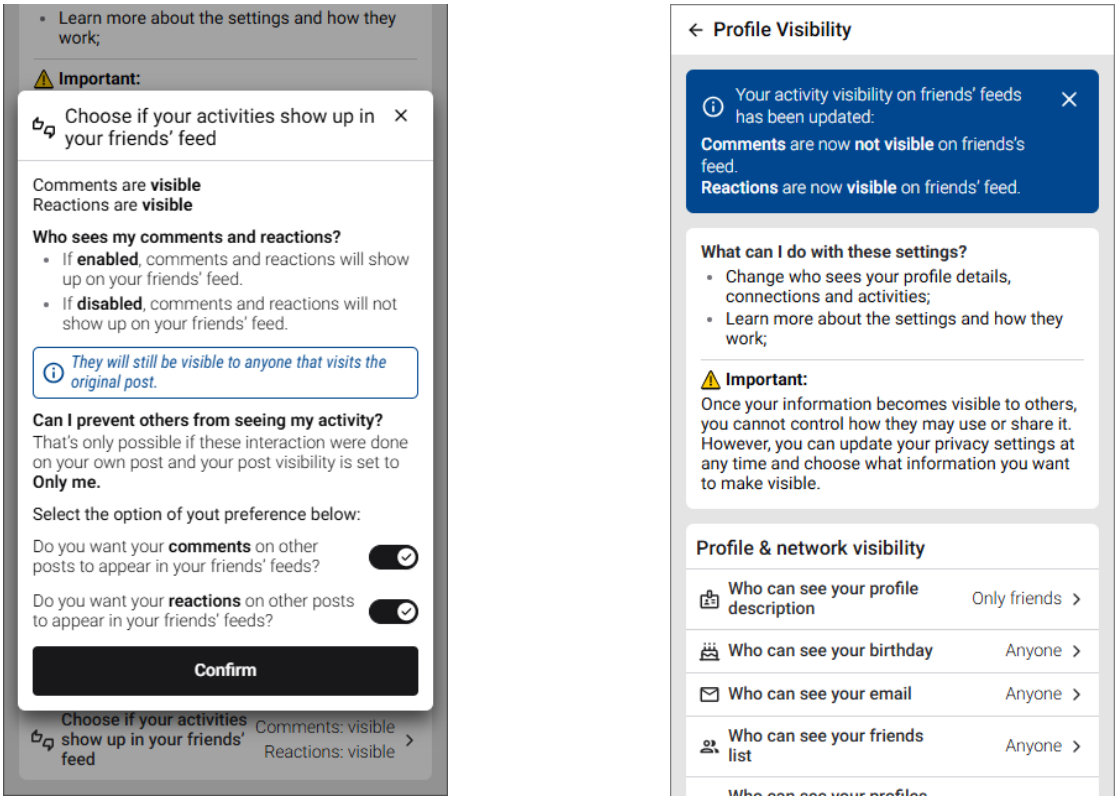
(a) (b)

Figure 11. *Choose if your activities show up in your friends' feed* (a) setting and (b) feedback, PH version.

the current status, followed by an explanation of the setting or its functionality, and concludes with information on limitations and actionable instructions.

The initial notice has been reformulated into a guiding question, accompanied by two bullet points outlining the implications of enabling or disabling the setting, along with an information box (outlined in blue). The information box clarifies that these activities remain visible to anyone who visits the original post where the comment or reaction was made. This revision aims to make the information more engaging and easier for users to navigate, while maintaining neutrality by presenting all options with equal clarity and prominence, an approach guided by **RPH6.2**, which promotes enhanced navigability through clearly recognizable elements, and a balanced and unbiased design.

Additional information was introduced in the form of a question and supporting text to address potential doubts about whether it is possible to completely hide reactions and comments from others. The presentation of this clarification as a question is intended to capture attention and encourage DSs to read the content, supporting informed decisions. This aligns with **RPH1.3**, which advocates for clear, meaningful and actionable guidance,



(a) (b)

Figure 12. *Choose if your activities show up in your friends' feed* (a) setting and (b) feedback, RPH version.

placed with relevant context. It also aligns with **RPH9.1** as it makes the instructions more digestible for a broader range of users.

Despite the increased informational load, the design remains simple, both visually and textually. Bold and lighter text are used strategically to help DSs distinguish between different pieces of information and to highlight key terms, consequently enhancing readability and supporting navigation. This design decision aligns with **RPH6.2** and **RPH9.1**, which emphasize the importance of presenting information in a balanced and easily navigable format, that also accommodates varying levels of digital literacy and potential cognitive disabilities. Additionally, we highlight the addition of icons to the toggle elements to make their status more immediately recognizable, as relying on color alone may not be sufficient to convey whether the setting is enabled or disabled. This design choice is once again guided **RPH6.2** and **RPH9.1**.

To conclude this step, we evaluate the new design against the acceptance criteria, following the same approach used in the previous settings. The evaluation is shown in Table 17.

Table 17. *Choose if your reactions or comments show up on your friends' feed* setting evaluation with AC.

DP1AC1. Are all choices presented with equal prominence? (i.e., Are they displayed in a way to allow users to perceive them as equally relevant?)
Yes. The design presents both privacy actions in the same way, only changing the functionality to which it refers. Information about these functionalities is also the same.
DP2AC1. Is the privacy-related choice or information presented accurately and comprehensible to users of different levels of expertise?
Yes. The information provided accurately reflects the functionality. Information that relates to how the privacy option work is placed at the beginning, where users are expected to be most attentive.
DP3AC1. Is the path to privacy-related mechanisms easy to navigate to?
Yes. We enhanced the instructions explaining how the mechanism works to make it easier to learn, and used contrast in the design to support intuitive navigation through this information.
DP3AC2. Are actionable privacy mechanisms (e.g., privacy settings) easily recognizable and intuitive to use?
Yes. The privacy mechanisms were slightly enhanced with the placement of icons in the toggles, to indicate the current status.
DP6AC1. Are the consequences of privacy choices clearly communicated to the user during or after a decision-making process, through real-time feedback or confirmation?
Yes. As soon as the DS confirm their changes, they are met with a feedback dialog reminding them of the change. Since it allows two actions, both changes are shown at once.
DP6AC2. Is information about the implications of users' privacy choices easy to find?
Yes. The information presented before the privacy action explains the consequences of each option. Additionally, we inform users about the setting's limitations, as they might expect to have full control over the visibility of comments and reactions across all posts.
DP9AC1. Is privacy guidance provided based on the user's current situation (e.g., location, device type, or task being performed) and interaction context (e.g., signing up for a service vs. sharing a photo)?
Yes. Although this setting enables two privacy actions they work the same way, which allows us to provide the instructions in a generalized but still comprehensible manner.

5 Validation

In this section, we describe and discuss the validation of the design principles by experts and the RPHs by an experiment conducted with end-users. Both validations are part of **Step 3. Approach evaluation**, and the latter is the last step in the methodological process, **Step 4. Validate the responsible privacy solution**.

5.1 Validation of the design principles via experts

To validate the design principles, we contacted experts in the privacy domain, such as those with experience in privacy engineering and usable privacy. These experts were invited to complete a structured questionnaire designed to evaluate the **clarity** and **applicability** of each design principle, the **validity** of each acceptance criteria, and the overall **completeness** of the proposed solution.

The questionnaire, available in the Appendix A, was organized as follows: starting with a brief introduction presenting the thesis topic, research context, and objectives. Followed by an explanation of the evaluation goals, presentation of the Likert-scale questions, and instructions on how to respond to these questions. Then, they were asked to assess each design principle and its associated AC, which was presented in a table. The table included a dedicated feedback space for comments on specific principles or criteria. Finally, it concluded with an open-ended question addressing the overall completeness of the design principles, along with an additional space for any final feedback.

The Likert-scale questions used a 1 to 5 scale, in which scores closer to 1 represent more negative evaluations and scores closer to 5 indicate more positive assessments. To determine whether a DP or an AC is considered satisfactory, we defined three evaluation categories based on score ranges. Each category is visually distinguished by a color, which will later be used to help interpret the experts' feedback at a glance. Table 18 summarizes these categories, their corresponding score ranges, and the associated evaluation meaning.

Table 18. Evaluation categories for Likert-scale questions.

Color	Score Range	Evaluation
Green	≥ 4	Positive evaluation — the principle or criterion is considered clear, applicable, or valid.
Yellow	≥ 3 and < 4	Neutral to slightly positive — the principle or criterion might require some minor adjustments.
Red	< 3	Negative evaluation — the principle or criterion might need substantial revision.

Notice that the evaluation meaning is not absolute. For this reason, we deliberately use the term "might" in the descriptions for the yellow and red categories, as a score

alone cannot definitively indicate whether a revision is needed. We interpret these quantitative scores in conjunction with the qualitative feedback provided by the experts, when available. In cases where no explanatory feedback is given, it is not possible for us to determine the specific concerns or gaps perceived by the expert. Furthermore, considering that the experts did not have access to a detailed demonstration of the practical applicability of the design principles and criteria, we assume that this lack of additional context could have contributed to lower scores in some instances.

Three privacy experts responded to the questionnaire. After reviewing their evaluations we noticed that one of the experts misinterpreted the purpose of the design principles. For that reason, we decided not to consider their evaluation in this validation step. Moving forward, we refer to the remaining evaluations as that of *Expert A* and *Expert B*, to facilitate referencing.

We calculated the average scores from their responses to the questionnaire and summarized them in Table 19. The average scores are presented using the color categories defined in Table 18, indicating their corresponding evaluation. The individual scores are available in Appendix B. In the following discussion, we analyze the lower scores in detail, taking into consideration the presence or absence of feedback. When feedback was given, we highlight the specific gaps or areas for improvement identified by the experts to better understand and address the underlying concerns.

Table 19. Average scores from expert evaluations on the Likert-scale questions.

	Q1. Clarity	**Q2. Applicability**	**Q3. Validity**
DP1	5.0	3.5	-
DP1AC1	-	-	3.0
DP2	5.0	3.5	-
DP2AC1	-	-	3.5
DP3	5.0	4.5	-
DP3AC1	-	-	3.0
DP3AC2	-	-	4.0
DP4	4.5	4.0	-
DP4AC1	-	-	3.0
DP5	3.5	3.5	-
DP5AC1	-	-	3.5
DP6	4.5	4.0	-
DP6AC1	-	-	4.0
DP6AC2	-	-	4.0
DP7	4.5	4.0	-
DP7AC1	-	-	3.5
DP8	4.5	4.0	-

DP8AC1	-	-	4.0
DP9	3.5	2.5	-
DP9AC1	-	-	3.0
DP10	5.0	4.0	-
DP10AC1	-	-	4.0
DP11	5.0	3.5	-
DP11AC1	-	-	3.5

We begin by discussing the scores in the lowest evaluation category, highlighted in red. Starting with the applicability criteria (**Q2**), the average score of the principle **DP9. Situation-aware** fell below 3.0. To better understand the reasons behind this low score, we examined the individual feedback provided by the experts. *Expert A* assigned this principle a score of 2.0 but did not provide any specific justification, which makes it difficult to understand the rationale behind considering the principle "not very applicable." In contrast, *Expert B* assigned a score of 3.0 and noted that the distinction between this principle and **DP8. Context-aware** could be further clarified to enhance its applicability.

Considering that the lowest score was not paired with any feedback, we weighed more on the score provided by *Expert B*, indicating the need of some minor adjustments. And, based on their feedback we believe that an example of the applicability of both **DP8. Context-aware** and **DP9. Situation-aware** could suffice to help practitioners distinguish and implement them. As mentioned previously, the experts did not have access to the demonstration part of this thesis, only a brief explanation about the topic (see Appendix A).

In regard to the scores categorized as neutral to slightly positive (highlighted in yellow), we identified a few principles and AC in this group. For **DP1. Neutral** and **DP1AC1**, the applicability (**Q2**) and validity (**Q3**) criteria, respectively, received scores within this range. *Expert B* evaluated these elements positively but suggested replacing the word "displayed" with "designed" in **DP1AC1**'s description, as the former may be somewhat limiting. *Expert A* assigned a score of 3.0 to the applicability of **DP1. Neutral** and a score of 2.0 to the validity of **DP1AC1**, explaining that there might be a cultural gap in privacy awareness among professionals.

Expert A argues that there might have instances in which developers may lack sufficient knowledge of privacy issues to effectively apply and evaluate these principles. However, we argue that a lack of privacy knowledge among professionals represents a limitation in how they might interpret and apply the design principle, rather than a flaw in the principle itself. The way in which a designer or developer implements the principle can vary significantly depending on their context, expertise, and experience.

The next principle and AC, **DP2. Honesty and Clarity** and **DP2AC1**, also fell into this category based on the same criteria as the previous ones (i.e., **Q2: Applicability** and

Q3: Validity). *Expert A* assigned a score of 3.0 to both but did not provide any feedback to justify these ratings. In contrast, *Expert B*, scored both elements positively with a 4.0 and offered constructive suggestions for refinement. These suggestions were minor changes, including mentioning *"users with varying levels of digital skill"*, and re-visiting the phrasing of *"presented accurately and comprehensibly"* to *"presented accurately, yet understandably"*, in **DP2AC1**'s description.

Although the first suggestion aims to emphasize inclusivity and accessibility in **DP2. Honesty and Clarity**, we considered its implementation unnecessary in this context, as we already address this concern explicitly in a dedicated principle, namely **DP10. Accessible and inclusive**.

Next in this category, we have **DP3AC1**, **DP4AC1**, **DP5. Benefit-risk balance**, and **DP5AC1**. Regarding **DP3AC1**, *Expert B* assigned a positive score of 4.0 but did not provide any suggestions for improvement. In contrast, *Expert A* gave it a lower score of 2.0, explaining that assessing users' learning is more challenging than evaluating their perception of usability. We acknowledge that measuring learning outcomes can be considerably more complex than assessing ease of use or perceived usability. Since this AC was originally intended to help designers verify the ease of navigation paths, we decided to incorporate *Expert A*'s feedback and refine its focus accordingly. See Table 20

Table 20. DP3AC1 before and after evaluation feedback.

Original version	**Refined version**
DP3AC1. Is the path to privacy-related mechanisms easy to learn and intuitive to use?	**DP3AC1**. Is the path to privacy-related mechanisms easy to navigate to?

In the case of **DP4AC1**, both experts assigned a score of 3.0 but did not provide any justification. However, considering that this principle and its AC focus on mitigating dark patterns of coercive nature (e.g., time limited offers, exploiting fear of missing out and bandwagon effect), which are often subjective and context-dependent, makes them inherently difficult to assess objectively. Consequently, a cautious or neutral score may reflect the challenges in evaluating the absence of such tactics rather than a perceived weakness in the criteria itself. Thus, we make no changes to this element.

DP5. Benefit-risk balance and **DP5AC1** both received an average score of 3.5 across all criteria (i.e., **Q1: Clarity**, **Q2: Applicability**, and **Q3: Validity**). *Expert A* assigned a score of 2.0 to the principle's clarity and applicability and 3.0 to the AC but did not provide any feedback to explain these ratings. Consequently, we place greater emphasis on the evaluation from *Expert B*, who rated the principle 5.0 for both clarity and applicability and 4.0 for the AC's validity, while providing a suggestion to improve these elements. They suggest refining the AC by including concrete examples to clarify the *"given action"* the AC refers to, which they considered somewhat vague.

Including examples directly in the descriptions of principles and ACs can be challenging, as it may make them overly detailed and difficult to navigate. We argue that, since the suggestion focuses on providing examples to guide designers, this can be effectively addressed through a dedicated demonstration of the principles and ACs in practice. Moreover, this approach can enhance designers' understanding and improve the overall performance of this principle and AC across all evaluation criteria.

Moving on to the last group of principles and ACs evaluated in this category, we have **DP7AC1**, **DP9. Situation-aware**, **DP9AC1**, **DP11. Regulation compliant**, and **DP11AC1**. These elements were placed in the neutral to slightly positive category mainly due to the lower scores assigned by *Expert A*, who did not provide any feedback to justify these evaluations. In contrast, *Expert B* assigned scores of 4.0 or higher to all of these elements and offered suggestions for some of them. The feedback provided for **DP9AC1** was already discussed when addressing the lowest score category.

Regarding **DP11. Regulation compliant**, *Expert B* emphasized the importance of evaluating this principle in relation to the other design principles to avoid mere "checkbox compliance." This comment does not suggest a direct change to the principle itself but rather serves as guidance to encourage designers to consider regulatory compliance as part of a broader, integrated approach to responsible privacy.

Since the principles and criteria that received positive evaluations are considered satisfactory, we will not discuss them further in this section. Instead, we now turn to the fourth and final question included in the questionnaire: an open-ended question focusing on the completeness of the overall set of design principles. The question was formulated as follows: *"Do the current design principles, collectively, cover necessary aspects of responsible privacy? Are they adaptable across different privacy contexts (e.g., social media, online banking, healthcare) and mechanisms (e.g., privacy settings and policies)?"*

It is important to note that the phrasing of this question naturally invites a yes-or-no response. For this reason, we explicitly encouraged the experts to elaborate on any perceived gaps and provide suggestions for improvement (as detailed in the final page of the Appendix A). This approach aimed to capture better insights and actionable feedback beyond a simple affirmative or negative answer.

Expert A's response did not provide a clear answer to this question, instead they made an observation about the relationship between compliance and ethics, noting that regulatory compliance does not necessarily equate to ethical practice. We suspect that this feedback was actually aimed at **DP11. Regulatory compliant**, which is the last principle presented prior to this open-ended question. Thus we cannot draw any insights regarding the completeness from this expert's response. In *Expert B*'s case, the response is more straightforward. They believe that the solution offers good coverage of distinct necessary aspects of responsible privacy, and suggest that we emphasize the need to consider these principles and ACs during the Privacy by Design cycle.

5.2 Validation of the responsible privacy solution via end-users

In this section, we perform **Step 4** of the methodological process: *Validate the responsible privacy solution*. To achieve this, we conducted an A/B test with potential end-users, comparing two design versions: a baseline interface created using existing PHs, and an enhanced interface developed using our proposed RPHs. The RPHs were systematically derived from the original PHs following our outlined approach.

We utilize the design developed previously for *Profile Visibility* settings, with the addition of the other settings from this group, not shown in Section 4. These settings design is available in the Appendix C. The interfaces were designed using Figma[4], a digital tool for design and prototyping. The design was then implemented using NextJS, ChakraUI and Typescript and deployed using Vercel[5].

The validation is structured around three components: defining the objectives, outlining the method (i.e., how the experiment will be conducted), and analyzing the results. For clarity, each component is presented in its own subsection.

5.2.1 Objectives and criteria

The goal of responsible privacy heuristics is to inform a design that empowers autonomy and informed decision. Keeping that in mind, this experiment aims to verify whether the RPH version fosters user autonomy and enables a more informed decision without compromising the usability. To do so, we adopt and evaluate the following criteria:

- **Perceived usability**: Measures how easy or difficult participants found the interface to use, ensuring that the RPH version does not fall behind compared to the PH version.
- **Perceived informed decision**: Assesses whether users felt they had sufficient information to make an informed privacy choice.
- **Informed decision**: Compares between the action of the user with the correct reference potential action.
- **Perceived autonomy**: Measures whether participants felt any pressure or influence from the interface during the decision-making.
- **Perceived consequence-awareness**: Measures whether participants felt that the interface helped them understand possible consequences/risks.
- **Decision awareness**: Cares for evidence that the user engaged with informative elements as inferred from interactions such as hovers, clicks, or dwell time.

[4]https://www.figma.com/
[5]https://vercel.com/

5.2.2 Methodology

We invited 14 participants with diverse backgrounds and varying levels of familiarity with social media and privacy settings to take part in the experiment. We noticed that one of the participants seemed to struggle with the interface due to language barrier so we did not consider their results. Another participant failed to pass the quality control question: *"Do you think using social media may negatively affect your privacy?"*. To which only those that responded with *Yes* could move forward with the experiment. The remaining 12 participants' demographics are summarized in Table 21.

Table 21. Participants' background information.

#	Age	Gender	Occupation	Education	Familiarity privacy settings
1	26-35	Man	Full-time employed	Bachelor's degree	More than 10 years
2	26-35	Woman	Full-time employed	Bachelor's degree	More than 10 years
3	26-35	Woman	Homemaker	Master's degree	More than 10 years
4	26-35	Woman	Full-time employed	Master's degree	More than 10 years
5	18-25	Woman	Student	High school	Between 5 and 10 years
6	26-35	Woman	Student	Master's degree	More than 10 years
7	26-35	Woman	Part-time employed	Bachelor's degree	More than 10 years
8	26-35	Man	Full-time employed	Bachelor's degree	More than 10 years
9	36-45	Man	Full-time employed	Master's degree	More than 10 years
10	36-45	Woman	Unemployed	Bachelor's degree	More than 10 years
11	18-25	Man	Student	Bachelor's degree	Between 5 and 10 years
12	26-35	Man	Full-time employed	High school	More than 10 years

The experiments were moderated, with the majority of them being conducted remotely, while a few were conducted in person. Each person interacted with only one version of the design. Nevertheless, the protocol followed by all is described bellow:

1. **Briefing**: Participant is given access to the experiment's Google Form (see Appendix D), in which they can read about the purpose of the experiment, what data will be collected and how (e.g., demographic information, screen recording, and post-task questionnaire). We let the participant read the instructions at their own pace and respond to any questions they might have. The duration of the experiment was already mentioned to the participant during the recruitment, but during the briefing, we emphasize that they can take their time with the experiment. Finally, before proceeding to the next section of the forms, the participant is asked whether they believe that social media affects their privacy (only those that respond *"Yes"* proceed). Informed consent is obtained before starting.

2. **Demographic information questionnaire**: Participant fill out their age range, gender, occupation, education, and familiarity with social media.

3. **Scenario**: Participant reads the scenario and lets the researcher know when they are ready to start. Then, the researcher provides the link to the application and instructs the participant to split the screen between the application and the scenario.

4. **Interaction and recording**: The researcher sets up the screen recording software (OBS Studio[6] was used for remote experiments and Mac's built-in screen recorder for in person experiments). As soon as the recording began, the participant is informed of which version they are going to interact with and get started with the task. When satisfied with their settings, the participant informs the researcher, who stops the recording.

5. **Post-task questionnaire**: Participant is instructed to go back to the Google Forms, to the next section after the scenario, in which they respond to the post-task questionnaire. After submitting their answers, the researcher will ask some additional questions to understand a certain choice, made during the interaction with the settings, or the reasoning behind some of the participant's response to the questionnaire.

The steps we expected the participant to take during the application interaction (not shown to the participant) are outlined bellow:

Steps expected from the participant

1. Read the scenario and choose the version of the interface according to the researcher's instruction. **Version A** for PH interface and **Version B** for RPH interface);

[6]https://obsproject.com/

2. Navigate the *"Profile visibility settings"* interface and click on a setting;
3. Review the setting's instructions and decide which privacy option best matches their preference;
4. Select the desired privacy option and confirm their choice. If the pre-selected option already aligns with their preference, they may simply close the setting;
5. Repeat the process for all the other settings and let the researcher know when they are done;
6. Complete the questionnaire related to this scenario/setting.
7. Respond to additional questions the researcher might have;

5.2.3 Results

The purpose of this experiment is to compare the privacy-preserving behavior of the participants regarding the two interfaces designed using PH and RPH, based on the criteria defined at the beginning of this section. In addition to responses collected via Google Forms, we gathered supplementary background information from participants, including their area of expertise and their typical social media behavior (e.g., whether they are active or passive users, and whether they prefer sharing content privately or publicly).

During the design interaction and the post-task questionnaire, we noted participants' behaviors and decisions. After completing the questionnaire, they were invited to explain the reasoning behind certain choices and responses. These qualitative insights were used to interpret the results and provide further context to the comparative performance of the two design versions.

To present the results clearly, we first describe participant demographics, followed by an analysis of the privacy settings chosen during the interaction, and finally a comparison of the questionnaire responses. Quantitative results are presented first, with qualitative findings used to support and contextualize the interpretation. At the end of this section, we discuss the findings, compare how each design version performed regards the criteria defined, and respond to our third research question: *How can we validate the usability and effectiveness of the developed RPHs in the final privacy solution?*

Participants demographic results

To illustrate the demographic distribution of participants, we present four graphs showing their age range, gender, education level, and familiarity with social media (Figure 13).

The sample consisted of 12 participants: 7 women and 5 men, aged between 18 and 45, with most falling in the 26–35 age range. The two youngest participants (18–25)

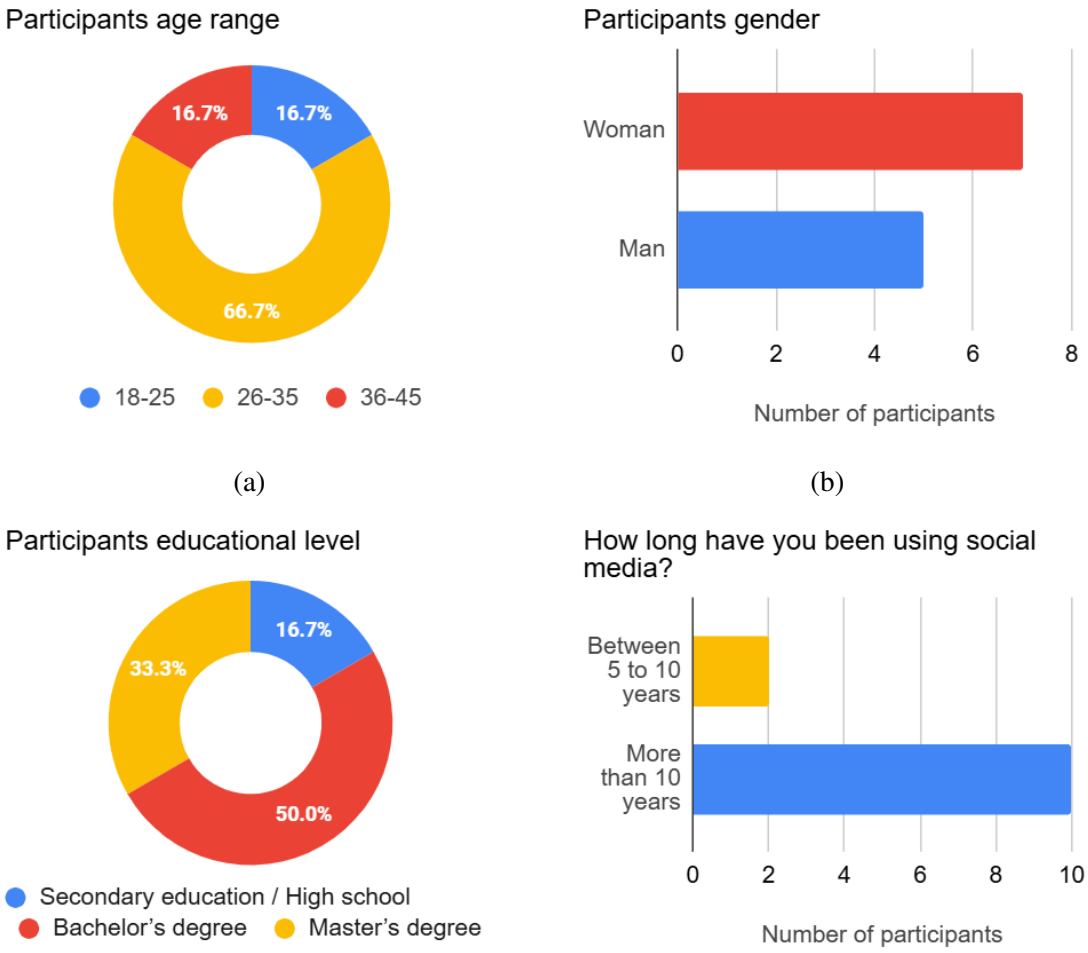


Figure 13. Participants demographic information (age range, gender, educational level, social media familiarity).

were also the only ones who had used social media for less than 10 years. One of them had completed secondary education, while the other held a bachelor's degree. In terms of educational background, half of the participants had a bachelor's degree and one-third held a master's degree.

We also collected information about participants' occupations and areas of expertise (Figure 14). Half of the participants were full-time employees, including four software engineers, one digital product designer, and one production machine operator. A quarter of the participants were students, pursuing degrees in cinema (bachelor's), physics (master's), and chemistry (PhD). The remaining participants included a home-

maker, a part-time worker, and an unemployed individual, with backgrounds in software engineering, architecture, and secretarial work, respectively.

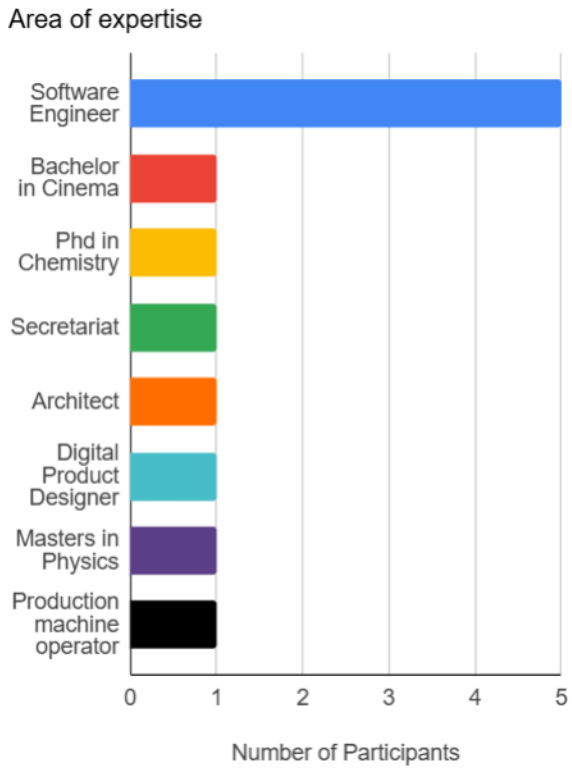


Figure 14. Participants area of expertise.

The final background information we collected concerned the participants' social media behavior. We asked them to describe what type of user they consider themselves to be on social media, how they interact with social media content, and whether they share personal posts. Based on their answers we categorized them as either **active** or **passive**. Although most participants described themselves as **passive**, and often conservative users, we considered additional behavioral indicators to better assess whether they truly fit that profile. A brief explanation of these profiles is provided below.

An active user is someone who regularly engages with the platform by creating and sharing content, whether publicly or privately, and frequently interacts with others by commenting on or reacting to posts. In contrast, **a passive user**, typically, uses social media to consume content with minimal interaction. They rarely engage with posts and, if they share content at all, they tend to do so only privately (e.g., via direct messages).

We found that four out of six participants who interacted with the PH version fit the

active user description, while only one participant in the RPH version did. Additionally, all **active** users were women, all men were **passive** users, and two women also identified as **passive** users. It is important to note that even among the **active** users, none described themselves as frequent posters or as individuals who regularly share their daily lives. Overall, based on their self-reports, participants appeared to maintain a relatively conservative relationship with social media.

When assigning participants to the design versions, we took most demographic characteristics into account, with the exception of their area of expertise and social media behavior, given that these were collected at the end of the experiment. The assignment was primarily balanced according to age range and educational level. As a result, the RPH version group included four men and two women, while the PH version group consisted of one man and five women. We later discuss how this distribution may have influenced the results.

Participants privacy choices results

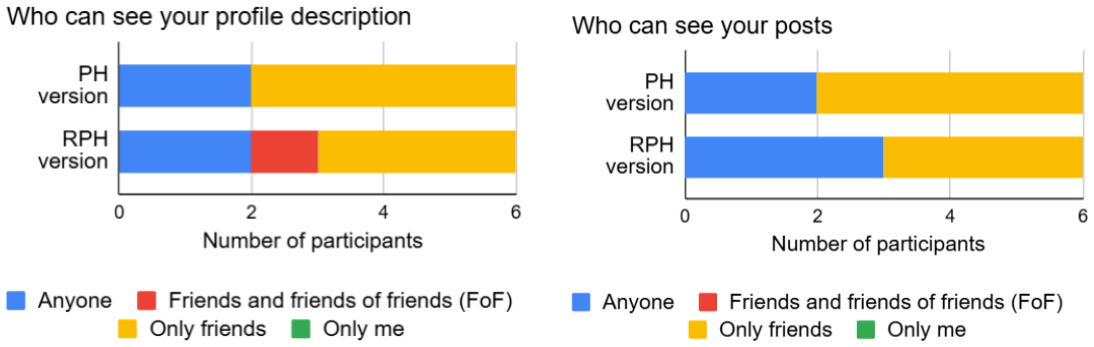
Regarding participants' privacy choices, we compare the two design versions using stacked bar graphs, which support the evaluation of the **Informed Decision** criteria. When paired with screen recording analyses, these results also contribute to assessing the **Decision Awareness** criteria. The latter focuses on user engagement with the interface, indicating whether participants took time to reflect on their choices or read the information presented.

Since the **Informed Decision** criteria aims to assess whether participants refrained from selecting the most public, privacy-invasive option (i.e., *Anyone*), we focus on investigating the reasoning behind those who selected this option. To do so, we review the experiment notes, including behavioral observations and participant justifications, and analyze screen recordings to examine how participants engaged with the interface. Based on this, we determine whether their choices can be considered informed.

The first settings analyzed are *"Who can see your profile description"* and *"Who can see your posts"*. These are grouped together because participants who chose **Anyone** for the first setting also chose it for the second, and provided similar justifications. Figures 15a and 15b show these results side by side.

Four participants, two from each design version, selected the privacy option **Anyone** for both settings, citing the "professional connection" aspect of the scenario. They considered it either necessary or beneficial to allow broad access to their workplace, education, and posts to increase their chances of expanding a professional network.

Two of these participants (one from PH and one from RPH) described themselves as primarily **passive** social media users who very rarely post. Nevertheless, within the given scenario, they preferred to keep posts public and instead manage visibility by being selective about what they shared. The other two, which we identified as **active** users,



(a) (b)

Figure 15. Results from (a) who can see your profile description and (b) who can see your posts.

emphasized the importance of publicly sharing posts to build a portfolio or reach a wider audience. An additional participant, also a **passive** user, selected **Anyone** only for post visibility, similarly justifying the choice as a way to showcase work-related content.

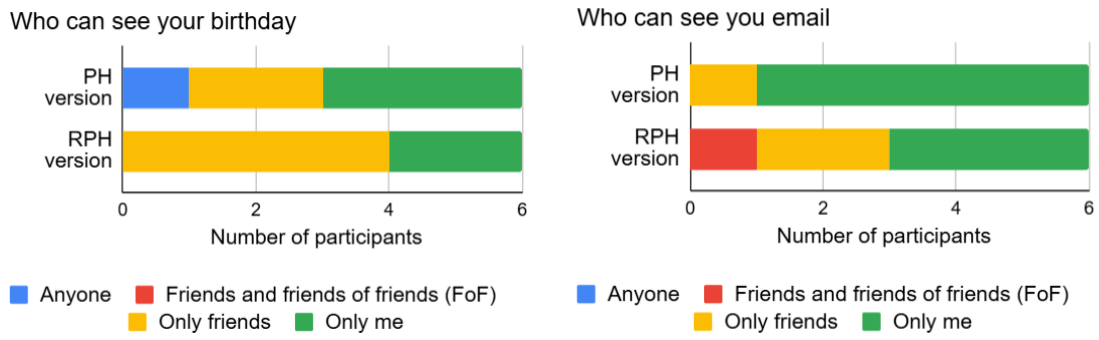
Regardless of the design version, participants motivated by professional networking goals appeared to view profile and post visibility as key tools. They opted to keep these elements public while exercising manual control over what content to include. This suggests that the interfaces were not particularly effective in steering users away from more public options. However, this does not necessarily indicate that their decisions were uninformed.

For instance, one participant using the RPH version, who ultimately selected **Anyone** for profile description, demonstrated clear deliberation. He initially selected a more privacy-preserving option **Only friends**, but after reviewing all settings and asking the moderator whether the profile fields (e.g., workplace, education, location) were mandatory, he returned to this setting and changed his selection to **Anyone**. Learning that these fields were optional allowed him to choose which information to exclude (e.g., location), making him feel more comfortable with a public visibility option. This example suggests that the participant made an informed decision, despite selecting a more public setting.

When looking through the screen recordings, we notice that the participant (RPH version) that chose **Anyone** for post visibility, but **Friends and friends of friends (FoF)** for profile description was influenced by the interface. This was inferred given that this participant seemed to spend extra time deliberating which option to select. In contrast, one of the participants (in PH version) that selected **Anyone** for both settings, did not seem to give much thought on his decision, by how quick they interacted with both settings.

With this context in mind, even though the participants made similar choices in both versions, the participants that interacted with the RPH version, seemed to be making these decision after some deliberation, compared to the PH version. This is also reinforced by the fact that the responsible design version seemed to have managed to retain more of users' attention to the information provided.

Next we show results for the settings *"Who can see your birthday"* and *"Who can see your email"*, illustrated by Figures 16a and 16b, respectively.



(a) (b)

Figure 16. Results from (a) who can see your birthday and (b) who can see your email.

For the birthday visibility setting, at first glance, the RPH version appears to have enabled more informed decisions compared to the PH version, as evidenced by the presence of a more public, privacy-invasive option selected in the PH interface. Further supporting this assumption is the fact that the participant who selected **Anyone** in the PH version explicitly stated that they did not perceive any risks in making their birthday visible to anyone.

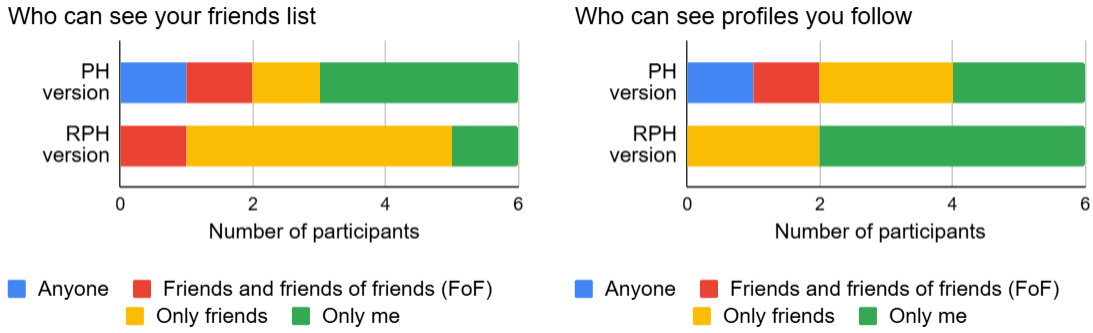
The screen recording did not provide much insight for this setting. Participants in the PH version actually spent a bit more time on average in this setting than the ones in RPH version. Since this was the second setting participants typically clicked on, and they already had seen what the privacy options were, the majority of them, regardless of the design version, would quickly selected their preferred option.

For the email visibility setting, participants who interacted with the PH version selected the most privacy-preserving option more frequently. However, we do not believe this outcome was influenced by the interface itself, as the PH version does not stimulate users to reflect on potential privacy risks, unlike the RPH version. In contrast, one participant from the RPH group, who ultimately selected the privacy option **Friends and friends of friends (FoF)**, had initially intended to choose **Anyone**. This change in decision was first inferred from their interaction patterns within the application and later confirmed during the session, when the participant explained to the moderator that the information presented in the setting led them to reconsider their initial choice.

The analysis of the screen recordings revealed that participants interacting with the RPH version spent approximately twice as much time, an average of 18 seconds, on this setting compared to those using the PH version, who spent around 8 seconds on average. While the increased duration may be partially explained by the additional information provided in the RPH interface, we believe that it also reflects deeper engagement with the decision. This interpretation is supported by cursor behavior: participants often used their cursor to follow along the text, allowing us to observe when they had finished reading and how they navigated through the setting. This pattern suggests that at least part of the additional time was spent reflecting on the choice, not just reading the content.

Based on this information, we infer that RPH version performed slightly better than PH version when it came to fostering informed decision making. We also observed improvement in user engagement with the information provided in the interface, which relates to the decision awareness criteria.

For the settings *"Who can see your friends list"* and *"Who can see profiles you follow"*, illustrated in Figures 17a and 17b, respectively, we observed that only the PH version had a participant choosing the privacy option **Anyone**, for both settings. He justified his choice by explaining that he prefers to shape his profile with a professional purpose in mind, rather than prioritizing his personal preferences when choosing which profiles to follow.



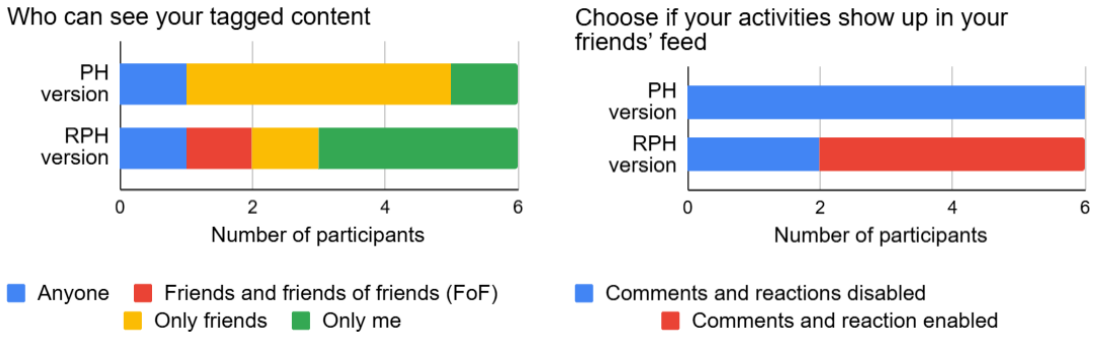
(a) (b)

Figure 17. Results from (a) who can see your friends list and (b) who can see profiles you follow.

Compared to the other settings, the differences between the designs of these two settings are minimal (see Figures 25 and 26), meaning that we expected participants from both versions to spend similar time interacting with them. That was indeed the case for the second setting, but not for the first. For the setting related to profiles you follow, in the RPH version, participants spent on average 14 seconds, while in the PH version they spent 8 seconds. This might indicate that the text insinuating that this feature (i.e.,

friends list) "reveals aspects of the user's social circle and connections" had some impact in their decisions.

The last pair of settings we analyze is *"Who can see your tagged content"* and *"Choose if your activities show up in your friends' feed"*, illustrated in Figures 18a and 18b.



(a) (b)

Figure 18. Results from (a) who can see your tagged content and (b) choose if your activities show up in your friends' feed.

For tagged content, we find that the outcome of both versions are a bit similar, with one participant from each design version choosing **Anyone**. The reasoning behind their choices further confirm that in both cases the decision was uninformed. The participant that interacted with the PH version stated that she assumed that the application would have another feature in which she can control who can tag her and she would be able to choose friends she trust (people that would not tag her on weird posts). As for the one that interacted with RPH version, he mentioned that he personally thinks this information is harmless.

Upon analyzing the screen recording, we noticed that the participant in the RPH version, that selected **Anyone**, was also the one that spent the longest time in this setting's interface compared to the other participants. This indicates that the RPH version failed to communicate the privacy risks or make the user reflect on possible risks.

As for the second setting, in which users can choose if their activities appear on friends' feeds, there are no privacy options that are completely public. So, to infer whether users made informed decisions, we rely on observing how they engaged with the interface and the notes made during the experiment.

We noticed that participants had some issues with this interface due to the amount of information provided in it, even in the PH version. This is highlighted by the open-ended question from the questionnaire, that evaluated perceived informed decision. However, one of the responses to the open-ended question also mentions another issue, in the PH version, related to comprehensible information about this setting.

When reviewing the screen recordings, we found that the majority of participants in both design versions engaged with the information presented in the interface, even though some appeared to find it somewhat overwhelming. The RPH version was able to retain users' attention despite being significantly more information-heavy than the PH version. Furthermore, given that the RPH version was slightly more effective in explaining how the last setting works, we regard it more positively than the PH version.

Nevertheless, both these settings point out a need to review and improve the RPH version of their interfaces, to foster informed decision making and improve decison awareness.

Post-task questionnaire results

For the post-task questionnaire results, we used stacked bar charts to compare responses across both versions, enabling a clear visual comparison. The first question, *"How easy or difficult was it to use this interface?"*, relates to the criteria **Perceived Usability**. Participants rated their experience on a 5-point Likert scale, from *5 – Very easy* to *1 – Very difficult*. The expectation was that the RPH version would perform as well as, or better than, the PH version, despite containing more information.

As shown in Figure 19, the RPH version performed equally to the PH version in terms of perceived usability, even with its added informational content.

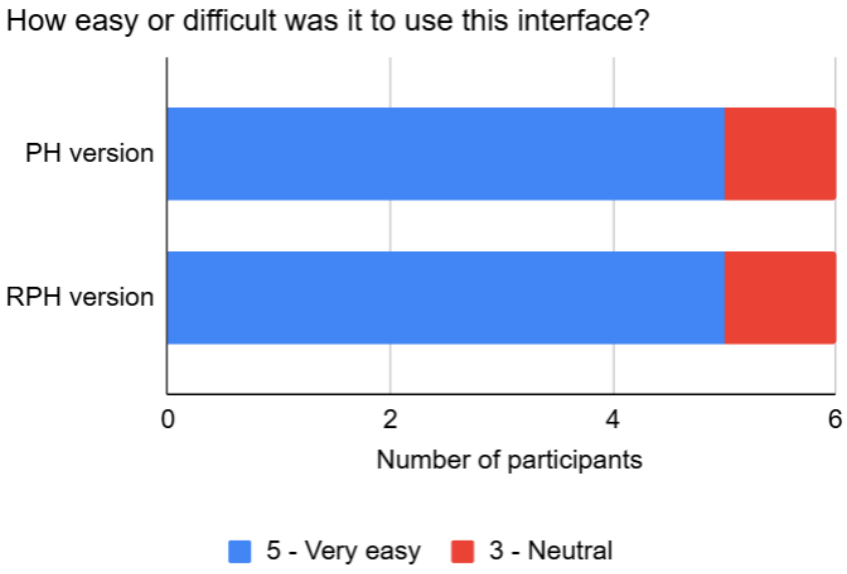


Figure 19. Perceived usability results post-task questionnaire.

To complement the quantitative findings, we analyzed the responses to the open-ended follow-up question: *"Were there any parts of the interface or information that made the decision difficult or confusing? Please describe."*. In which the responses were organized in Table 22 for easier reference. Two responses were omitted from the table as they did not relate directly to the interface.

Table 22. Perceived usability open-ended question responses.

Version	**Score**	**Response**
PH version	5-Very easy	Mentioned that the final setting "Choose if your activities…" contained too much text.
	3-Neutral	Found the same setting unclear regarding the extent of activity exposure to others.
RPH version	5-Very easy	Felt the final setting included too much information in a single window, which was somewhat overwhelming.
	5-Very easy	Mentioned that he was not sure how tag functionality works for this application, and assumed it works like Facebook / Instagram's.
	3-Very easy	Reported that the amount of on-screen text was confusing, and the split block at the bottom seemed easy to overlook.

Notably, even participants who rated usability as "very easy" commented on the interface being text-heavy. This suggests that excessive text may influence clarity or increase cognitive load, regardless of the overall perceived ease of use. In the PH version, one participant commented on unclear communication regarding the extent of activity visibility. In the RPH version, a participant expressed uncertainty about how the tag functionality would operate, since they did not have the opportunity to interact directly with the social media application itself.

The next criteria, **Perceived Informed Decision**, is also measured by one Likert-scale and one open-ended question. The Likert-scale question asked: *"How confident are you that the option you selected reflects your actual privacy preference?"*. Participants scored between *5-Confident* and *1-Not confident at all*, resulting in Figure 20. The figure shows that RPH version performed slightly better than PH the version.

For the open-ended question *"Did you feel that the privacy settings interface offered enough information for you to confidently decide who should see your profile details?*

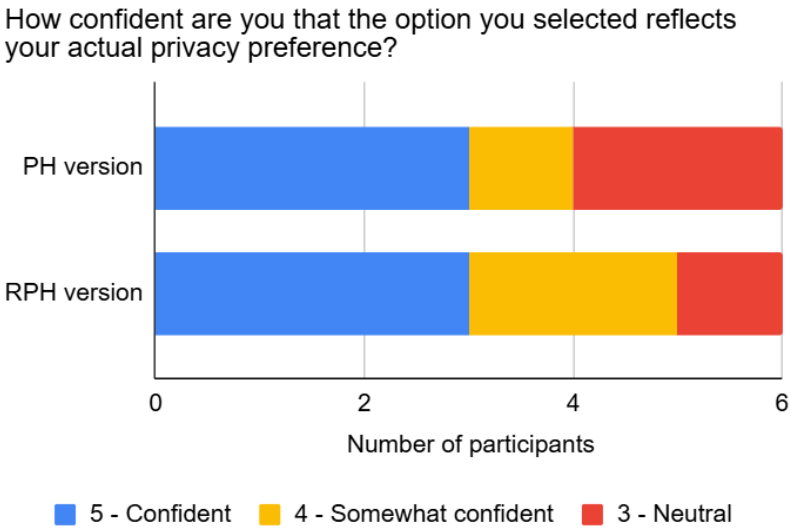


Figure 20. Perceived informed decision results post-task questionnaire.

If not, what was missing?", all participants responded positively, indicating that both interfaces provided sufficient information for informed decision-making. One participant from RPH version repeated his comment from the previous open-ended question (the one in the fourth row of Table 22). This might suggest that there is a need to improve the clarity of the tag setting's description or reflects a limitation of the experiment context, where the functionality could not be directly demonstrated.

As most participants did not elaborate further in their open-ended question responses, we followed up with those who rated the question as *3 - Neutral* to better understand their reasoning. In the PH version group, one participant indicated a general distrust in social media platforms, explaining that they do not believe privacy settings function as promised, an attitude stemming from skepticism toward platforms rather than this interface itself. Another participant from the same group mentioned they would have preferred to be able to customize options, such as setting visibility for different user groups separately. In the RPH version group, the neutral score was explained by a participant who stated they had no strong opinion on the options and were indifferent, as they rarely use social media.

The next criteria is **Perceived Autonomy**, which was measured by the question *"To what extent did you feel pressured by the information in the privacy setting interface to choose or not choose a specific option?"*. Participants had to choose scores between *5 - Not at all pressured (I felt completely free to make any choice without influence)* and *1 - Very strongly pressured (I felt I had little to no freedom to choose anything else)*. Figure 21 shows the comparison between both versions.

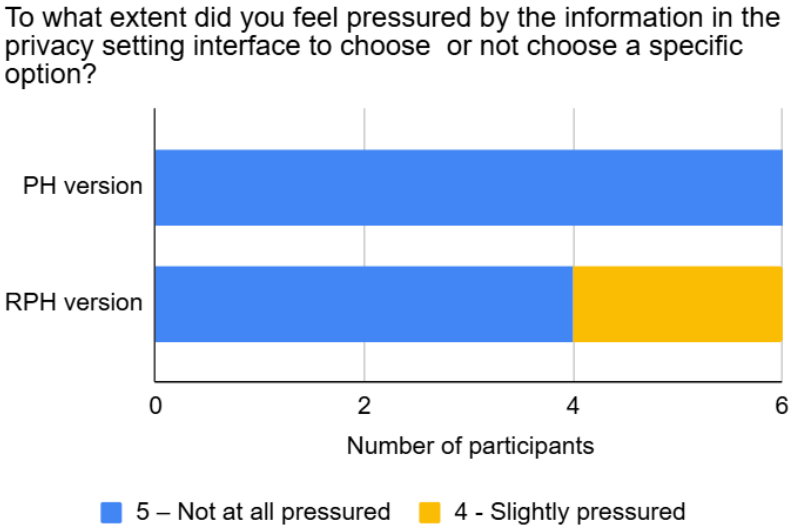


Figure 21. Perceived autonomy results post-task questionnaire.

It is noticeable that participants in the RPH version felt a little bit of *"pressure"*, compared to the PH version. More specifically, they noticed a small amount of influence. This is not necessarily something negative, since the RPH version does contain additional information aiming to make them carefully consider their choices. In fact, it indicated a positive impact given that one of the participants who scored *4-Slightly pressured* confirmed that they initially thought about picking privacy option **Anyone**, for email visibility, but after reading the information box they changed their mind, choosing a slightly more privacy-friendly option, **Friends and friends of friends (FoF)**. The other participant with the same score also mentioned that they noticed these pieces of information across the settings, which made them think a bit more before making a decision.

The final criteria addressed by the post-task questionnaire is **Perceived Consequence Awareness**, measured by the question *"To what extent did the information provided in the privacy settings interface help you understand the potential consequences of your choices (e.g., privacy risks)?"*. The participants choose a score between *5 - Completely (I clearly understood the potential consequences of each choice)* and *1 - Not at all (I didn't feel more informed about potential consequences)*. The results, illustrated in Figure 22, show that the RPH version scored more positively than the PH version.

We observed that participants often answered this question based on their personal perceptions of privacy risks, rather than directly referencing cues or information presented within the interface. This was confirmed during the experiment, either through participants' think-aloud comments or when prompted by the moderator. Such behavior

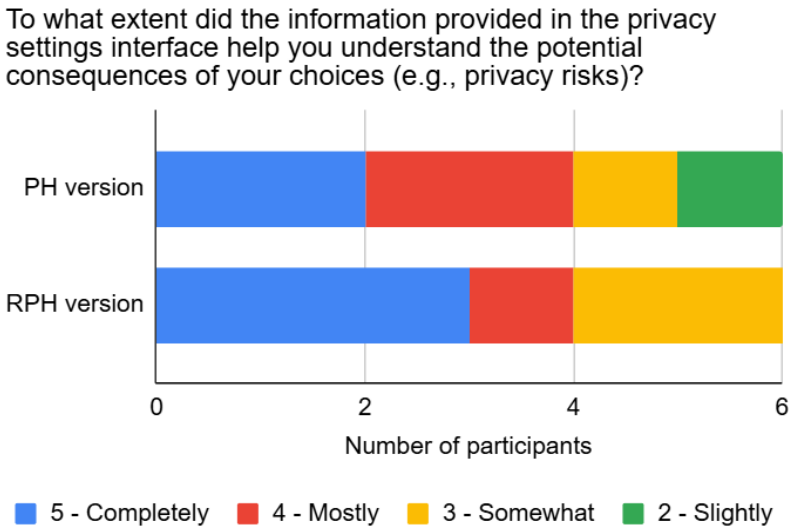


Figure 22. Perceived consequence awareness results post-task questionnaire.

helps explain why the PH version, despite lacking any explicit indicators of risk or prompts for reflection, still received scores of *3 – Somewhat (I understood some of the possible consequences)* or higher. Participants may have drawn on prior knowledge or beliefs about social media privacy, rather than the design itself.

In that context, we also acknowledge that selecting *5 - Completely (I clearly understood the potential consequences of each choice)* or *4 - Mostly (The information gave me a good understanding)* for the RPH version may reflect a slight overestimation. While the RPH design does introduce informative elements to raise awareness of privacy risks, it is not exhaustive. We intentionally avoided providing an extensive or overly explicit list of potential risks, as doing so could result in an interface that is either overwhelming or overly alarming. Both of which may compromise user autonomy and impair their ability to make a balanced, informed decision.

Nonetheless, at least three participants who interacted with the RPH version explicitly noted the presence of information that encouraged them to reflect on privacy risks. One participant stated that this information led them to choose a more conservative privacy setting. Another participant, who selected *3 – Somewhat (I understood some of the possible consequences)* acknowledged the inclusion of risk-related information but in a general and non-specific way, more like a "soft warning" rather than a detailed alert.

Discussion of results

We found that half of the participants, three from each design version, selected the

most public privacy option (**Anyone**) for at least one setting. Participants using the RPH version selected this option 25% less frequently than those using the PH version. Occupation, area of expertise and educational level did not appear to influence the choices. Gender and social media profile also did not appear to influence these choices, given that the group was evenly split between men and women, and between **active** and **passive** users. Notably, the participant who selected the public option most frequently (in four settings) was a male, passive user who interacted with the PH version.

We summarize the evaluation according to the defined criteria:

- **Perceived Usability**: Both designs were perceived as usable, though participants in both versions noted the presence of too much text. This means that using RPH did not overwhelm or obstruct users from completing their tasks. Since the settings fit within a single screen view, alternative formats (e.g., audio or video) should be explored to improve accessibility and reduce cognitive load.
- **Perceived Informed Decision**: The RPH version slightly outperformed PH, with more users acknowledging awareness of privacy implications.
- **Perceived Autonomy**: Some participants felt more influenced by the RPH version. Two users mentioned that the privacy risk cues made them reflect more on their decisions, with one of them changing their original choice (**Anyone** to **Friends and friends of Friends (FoF)**) after some deliberation.
- **Perceived Consequence Awareness**: RPH showed better performance, with three participants explicitly mentioning noticing the information related to privacy risks. As mentioned on the previous criteria, one participant even changed his initial choice after becoming more aware of possible consequences
- **Informed Decision**: RPH was more effective in supporting informed decision-making, as it delivered richer information without losing user engagement and helped prevent at least one participant from selecting the least private option.
- **Decision Awareness**: RPH participants appeared to engage more deeply with the interface and the information it provided.

6 Threats to validity

Although this research provides a valuable contribution to the development of responsible privacy design, certain potential threats to validity must be acknowledged [22]:

Internal validity refers to how accurately the observed relationship in the study reflects a true cause-and-effect connection between the condition and the outcome. One concern is that participants might not behave naturally when adjusting their settings due to being observed. To mitigate this, we emphasized that they should act as they normally would and reassured them that they were not being personally evaluated. Another concern is potential bias from the experiment moderator. To minimize this, the moderator's role was strictly limited to answering participants' procedural questions unrelated to the interface itself, avoiding any guidance on specific choices or explanations about the settings' functionality, which was designed to be conveyed by the interface alone.

External validity concerns how well the findings can be generalized. In this research, the approach was demonstrated using one relatively simple and straightforward privacy mechanism based on a generic social media application. While relevant, more complex mechanisms like cookie management or privacy policies were not tested. Further research is needed to assess the approach's applicability across diverse privacy contexts and user groups. Additionally, the use of this approach in a real-world application, especially if it presents some deceptive pattern, could be of interest to validate how well this solution mitigates these patterns.

Conclusion validity reflects the degree to which the conclusions drawn from the study are reasonable. Given the exploratory nature of the validation and the limited participant diversity, the conclusions about the effectiveness and usability of the responsible privacy heuristics served to indicate a small but positive impact. More extensive studies with larger and more varied samples would strengthen the confidence in these findings.

7 Conclusion

The aim of this thesis was to develop an approach for designing responsible privacy heuristics (RPHs) to support designers in creating ethical and effective privacy mechanisms. Following the design science research methodology, the work resulted in 11 design principles, each accompanied by acceptance criteria, and a methodological process for applying them across a variety of privacy solutions.

The thesis addressed three research questions. First, through a review of literature on privacy heuristics, deceptive patterns, ethical principles, and ethical design, we defined what constitutes an RPH (RQ1). Second, we proposed a structured process for deriving RPHs from ethical principles and integrating them into the design of privacy solutions (RQ2). Finally, we validated the usability and effectiveness of RPHs through an end-user experiment comparing two version of an interface, one based on traditional privacy heuristics and the other on RPHs (RQ3).

Results showed that RPH-based designs maintained usability while slightly improving user awareness of privacy risks and supporting more informed decision-making, without reducing user autonomy.

The main contribution of this thesis is a practical approach for translating ethical principles and requirements into the creation and evaluation of responsible privacy heuristics, enabling practitioners to embed these principles into privacy design without compromising usability, and ultimately fostering more ethical and privacy-friendly systems. Given that the three RQs were properly answered, we advocate that this research has reached its objectives.

Future research should expand on the demonstration of the proposed approach by applying each of the design principles to a broader range of privacy solutions, ensuring that at least one concrete example is provided for each principle. This would help illustrate their versatility, offer clearer guidance for practitioners, potentially find gaps and ways to improve the approach. To broaden the range and complexity of the privacy solution, future research could focus on the application of the approach to privacy mechanisms, such as cookie consent banners, privacy policies, and other regulatory-driven disclosures. These mechanisms often involve intricate trade-offs between legal compliance, user comprehension, and business goals, which could help assess the robustness and adaptability of the produced responsible privacy heuristics.

References

[1] AHUJA, S., AND KUMAR, J. Conceptualizations of user autonomy within the normative evaluation of dark patterns. In *Ethics and Information Technology* (2022).

[2] BRIGNULL, H., LEISER, M., SANTOS, C., AND DOSHI, K. Deceptive patterns – user interfaces designed to trick you, April 2023.

[3] BÖSCH, C., ERB, B., KARGL, F., KOPP, H., AND PFATTHEICHER, S. Tales from the dark side: Privacy dark strategies and privacy dark patterns. *Proceedings on Privacy Enhancing Technologies 2016* (07 2016), 237–254.

[4] CARAGAY, E., XIONG, K., ZONG, J., AND JACKSON, D. Beyond dark patterns: A concept-based framework for ethical software design. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (2024), Association for Computing Machinery, pp. 1–16.

[5] CRONHOLM, S., AND GÖBEL, H. Guidelines supporting the formulation of design principles. In *Australasian Conference on Information Systems* (12 2018).

[6] D'OLIVEIRA, N., AND CUNHA, F. Brazilian general data protection law (lgpd): the relationship between information policy and information regime. *RDBCI Revista Digital de Biblioteconomia e Ciência da Informação 22* (08 2024).

[7] EUROPEAN PARLIAMENT AND COUNCIL OF THE EUROPEAN UNION. General data protection regulation (gdpr), 2016. Article 5.

[8] EUROPEAN UNION. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 04 2016. Official Journal of the European Communities.

[9] GHARIB, M. Privacy and informational self-determination through informed consent: The way forward. In *Computer Security. ESORICS 2021 International Workshops* (Cham, 2022), Springer International Publishing, pp. 171–184.

[10] GHARIB, M. Towards a heuristic model for usable privacy. In *Joint Proceedings of RCIS Workshops and Research Projects Track* (May 2024), pp. 1–10.

[11] GHARIB, M., GIORGINI, P., AND MYLOPOULOS, J. Copri v.2 — a core ontology for privacy requirements. *Data Knowledge Engineering 133* (2021), 101888.

[12] GUNAWAN, J., SANTOS, C., AND KAMARA, I. Redress for dark patterns privacy harms? a case study on consent interactions. In *Proceedings of the 2022 Symposium on Computer Science and Law* (2022), Association for Computing Machinery, p. 181–194.

[13] HERTWIG, R., AND PACHUR, T. *Heuristics, History of*. 12 2015, pp. 829–835.

[14] HJEIJ, M., AND VILKS, A. A brief history of heuristics: how did research on heuristics evolve? *Humanities and Social Sciences Communications 10* (2023).

[15] IWASE, H. Japan overview of the act on the protection of personal information. *European Data Protection Law Review* (2019).

[16] JACOS, D., AND MCDANIEL, T. A survey of user experience in usable security and privacy research. In *HCI for Cybersecurity, Privacy and Trust* (2022), Springer, Cham.

[17] KISSELBURGH, L., AND BEEVER, J. *The Ethics of Privacy in Research and Design: Principles, Practices, and Potential*. Springer International Publishing, 2022, pp. 395–426.

[18] KITKOWSKA, A. *The Hows and Whys of Dark Patterns: Categorizations and Privacy*. Springer International Publishing, 2023, pp. 173–198.

[19] KUNEVA, M. Roundtable on online data collection, targeting and profiling. Keynote Speech, March 2009.

[20] MARMION, V., BISHOP, F., MILLARD, D. E., AND STEVENAGE, S. V. The cognitive heuristics behind disclosure decisions. In *Social Informatics* (2017), Springer International Publishing, pp. 591–607.

[21] MATHUR, A., KSHIRSAGAR, M., AND MAYER, J. What makes a dark pattern... dark? design attributes, normative considerations, and measurement methods. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021), Association for Computing Machinery, pp. 173–198.

[22] MATTHAY, E. C., AND GLYMOUR, M. M. A graphical catalog of threats to validity: Linking social science with epidemiology. *Epidemiology 31*, 3 (2020), 376–384.

[23] MILDNER, T. *Mitigating dark patterns through responsible design - ethical design considerations for user-centred technologies*. PhD thesis, Universität Bremen, 2024.

[24] MÖLLER, F., GUGGENBERGER, T. M., AND OTTO, B. Towards a method for design principle development in information systems. In *Designing for Digital Transformation. Co-Creating Services with Citizens and Industry* (2020), Springer International Publishing, pp. 208–220.

[25] OFFERMANN, P., BLOM, S., SCHÖNHERR, M., AND BUB, U. Artifact types in information systems design science – a literature review. In *Global Perspectives on Design Science Research* (Berlin, Heidelberg, 2010), Springer Berlin Heidelberg, pp. 77–92.

[26] PARRILLI, D. M. *Defining a Privacy Ethical Framework for Service Design*. Springer Nature Switzerland, Cham, 2025, pp. 89–121.

[27] PATTAKOU, A., MAVROEIDI, A.-G., DIAMANTOPOULOU, V., KALLONIATIS, C., AND GRITZALIS, S. Towards the design of usable privacy by design methodologies. In *2018 IEEE 5th International Workshop on Evolving Security & Privacy Requirements Engineering (ESPRE)* (2018), IEEE, pp. 1–8.

[28] PEFFERS, K., TUUNANEN, T., ROTHENBERGER, M., AND CHATTERJEE, S. A design science research methodology for information systems research. *Journal of Management Information Systems 24* (01 2007), 45–77.

[29] POTEL-SAVILLE, M., AND DA ROCHA, M. From dark patterns to fair patterns? usable taxonomy to contribute solving the issue with countermeasures. In *Privacy Technologies and Policy* (2024), Springer Nature Switzerland, pp. 145–165.

[30] REIS, B., AND GHARIB, M. Towards an approach for designing responsible privacy heuristics. In *Proceedings of the RCIS 2025 Workshops* (2025), vol. 3987 of *CEUR Workshop Proceedings*, CEUR-WS.org. Workshop paper.

[31] RENAUD, K., AND ZIMMERMANN, V. Ethical guidelines for nudging in information security privacy. *International Journal of Human-Computer Studies 120* (2018), 22–35.

[32] SMUTS, H., WINTER, R., GERBER, A., AND VAN DER MERWE, A. "designing" design science research – a taxonomy for supporting study design decisions. In *The Transdisciplinary Reach of Design Science Research* (2022), Springer International Publishing, pp. 483–495.

[33] SPIEKERMANN, S., ACQUISTI, A., BÖHME, R., AND HUI, K.-L. The challenges of personal data markets and privacy. *Electronic Markets 25* (April 2015), 161–167.

[34] SUNDAR, S. S., KIM, J., ROSSON, M. B., AND MOLINA, M. D. Online privacy heuristics that predict information disclosure. In *Proceedings of the 2020 CHI*

Conference on Human Factors in Computing Systems (2020), Association for Computing Machinery, p. 1–12.

[35] ZIMMERMANN, V. *Privacy Nudges and Informed Consent? Challenges for Privacy Nudge Design*. Springer International Publishing, 2023, pp. 155–171.

A Expert Evaluation Questionnaire

Design principles for responsible privacy heuristics - an evaluation questionnaire

This questionnaire is part of a Master's thesis titled *"An Approach for Designing Responsible Privacy Heuristics".*

Context: Privacy compliance is a major concern for legal entities handling Personal Information (PI), as noncompliance leads to substantial fines. Regulations require these entities to implement privacy protection mechanisms (privacy solutions) and inform data subjects (DSs) about PI processing. However, DSs often struggle to understand relevant information and effectively use these mechanisms, leaving their privacy vulnerable. Privacy heuristics offer a potential solution by assisting users in making informed decisions. Yet, their design is complex, prone to bias, and, if done irresponsibly, may lead to unethical or manipulative outcomes. This research aims to address these challenges by developing an approach that offers design principles to guide the design and development of Responsible Privacy Heuristics (RPHs) for usable privacy-aware systems or solutions. These design principles have been formulated based on best practices in the responsible and ethical design area, taking into consideration available fair design patterns and trying our best to avoid the dark design patterns.

These design principles are intended to enhance traditional usability privacy heuristics, transforming them into RPHs. The acceptance criteria serve as a tool to assess whether a design resulting from an RPH correctly applies the principle.

The principles and their acceptance criteria (AC) are presented in Table 1. Given your experience, we seek your input to evaluate the **clarity** and **applicability** of these design principles, as well as the **validity** of their associated AC. We define these criteria as follows:

DP - Q1. Clarity: How easy is it to understand what this design principle is recommending from a design perspective?
1. Very difficult 2. Difficult 3. Neutral 4. Easy 5. Very easy

DP - Q2. Applicability: Can this principle be realistically applied in real-world design scenarios?
1. Not at all 2. Not very applicable 3. Neutral 4. Somehow applicable 5. Applicable

DPAC - Q3. Validity: Can this AC be applied, tested, and verified objectively by involved stakeholders (e.g., developers, testers) without ambiguity?
1. Not at all 2. Not sufficiently valid 3. Neutral 4. Somehow valid 5. Valid

Please evaluate each Design Principle (DP) and its related Acceptance Criteria (AC) by filling in the table below. Use the 1 to 5 scales provided in the questions above. You may use the Comments row to offer suggestions, note strengths or limitations, or share any relevant insights based on your expertise.

Table 1. Design principles and their acceptance criteria for responsible privacy heuristics

DP1. Neutral: A RPH should present information about privacy choices in a neutral and balanced manner, avoiding framing that could lead to biased or skewed decisions.	**Q1**	**Q2**
	1	1
DP1AC1. Are all choices presented with equal prominence? (i.e., Are they displayed in a way to allow users to perceive them as equally relevant?)	**Q3**	
	1	
Please use this space if you have any suggestions to improve or extend DP1 or DP1AC1		
DP2. Honesty and Clarity: A RPH should ensure that all information presented to users is truthful, clear, and easy to understand.	**Q1**	**Q2**
	1	1
DP2AC1. Is the privacy-related choice or information presented accurately and comprehensibly to users of different levels of expertise?	**Q3**	
	1	
Please use this space if you have any suggestions to improve or extend DP2 or DP2AC1		
DP3. Navigable and actionable privacy: A RPH should help users easily identify, understand, and act upon privacy-related information.	**Q1**	**Q2**
	1	1
DP3AC1. Is the path to privacy-related mechanisms easy to learn and intuitive to use?	**Q3**	
	1	
DP3AC2. Are actionable privacy mechanisms (e.g., privacy settings) easily recognizable and intuitive to use?	1	
Please use this space if you have any suggestions to improve or extend DP3, DP3AC1, or DP3AC2		
DP4. Pressure-free: A RPH should not impose time constraints, emotional manipulation, or other coercive tactics that pressure users into making privacy decisions.	**Q1**	**Q2**
	1	1
DP4AC1. Are users free to make privacy decisions without being subject to: time constraints; exclusive time-limited offers or other alleged financial gains in exchange for PI; coercive tactics exploiting emotional and social factors (e.g., guilt shaming, fear of missing out, and bandwagon effect)?	**Q3**	
	1	
Please use this space if you have any suggestions to improve or extend DP4 or DP4AC1		

DP5. Benefit-risk balance: A RPH should highlight user benefits and proactively minimize potential privacy risks.	**Q1**	**Q2**
	1	1
DP5AC1. Are the benefits and potential privacy risks associated with a given action clearly communicated?	**Q3**	
	1	
Please use this space if you have any suggestions to improve or extend DP5 or DP5AC1		
DP6. Consequences awareness: A RPH should provide feedback related to privacy choices, avoiding obscuring consequences, which could affect the users' decision-making.	**Q1**	**Q2**
	1	1
DP6AC1. Are the consequences of privacy choices clearly communicated to the user during or after a decision-making process, through real-time feedback or confirmation?	**Q3**	
	1	
DP6AC2. Is information about the implications of users' privacy choices easy to find?	1	
Please use this space if you have any suggestions to improve or extend DP6, DP6AC1 or DP6AC2		
DP7. Empowering: A RPH should support users to select privacy choices that align with their privacy requirements.	**Q1**	**Q2**
	1	1
DP7AC1. Does the privacy solution provide intuitive and customizable privacy options that enable users to control, correct, and retract their privacy choices in a way that aligns with their preferences and requirements?	**Q3**	
	1	
Please use this space if you have any suggestions to improve or extend DP7 or DP7AC1		
DP8. Context-aware: A RPH should help users assess privacy decisions in context, considering factors such as data sensitivity, purpose of collection and use, recipient identity, and potential risks.	**Q1**	**Q2**
	1	1
DP8AC1. Does the privacy solution provide users with context-specific information that allows them to assess the sensitivity and risks implied by the type of data being requested (e.g., health, financial, or personal data), the purpose for its use, and the identity of the recipients?	**Q3**	
	1	
Please use this space if you have any suggestions to improve or extend DP8 or DP8AC1		
DP9. Situation-aware: A RPH should adapt to different situations to provide relevant, meaningful, and actionable guidance.	**Q1**	**Q2**

	1	1
DP9AC1. Is privacy guidance provided based on the user's current situation (e.g., location, device type, or task being performed) and interaction context (e.g., signing up for a service vs. sharing a photo)?	**Q3**	
	1	
Please use this space if you have any suggestions to improve or extend DP9 or DP9AC1		
DP10: Accessible and inclusive: A RPH should ensure that users, regardless of their abilities or technical expertise, can understand and act upon privacy-related information.	**Q1**	**Q2**
	1	1
DP10AC1. Is privacy-related information provided in multiple formats (e.g., simple language, assistive technologies, alternative formats) to accommodate users' varying preferences, expertise, sensory needs, and disabilities?	**Q3**	
	1	
Please use this space if you have any suggestions to improve or extend DP10 or DP10AC1		
DP11. Regulation Compliant: A RPH must not encourage or lead to violating privacy legislation (e.g., purpose limitation, data minimization).	**Q1**	**Q2**
	1	1
DP11AC1. Are privacy-related choices and information compliant with the relevant privacy legislation (e.g., GDPR in Europe)?	**Q3**	
	1	
Please use this space if you have any suggestions to improve or extend DP11 or DP11AC1		

Q4. Completeness: Do the current design principles, collectively, cover necessary aspects of responsible privacy? Are they adaptable across different privacy contexts (e.g., social media, online banking, healthcare) and mechanisms (e.g., privacy settings and policies)?

Please use this space to answer Q4. Feel free to point out gaps and make suggestions.

Final Feedback: If you have any additional comments, suggestions, or concerns about this approach that were not covered in the previous questions, please share them here.

Please use this space for the final feedback.

Thank you for participating!
Beatriz Pontes Da Costa Reis

B Experts individual scores to Likert-scale questions

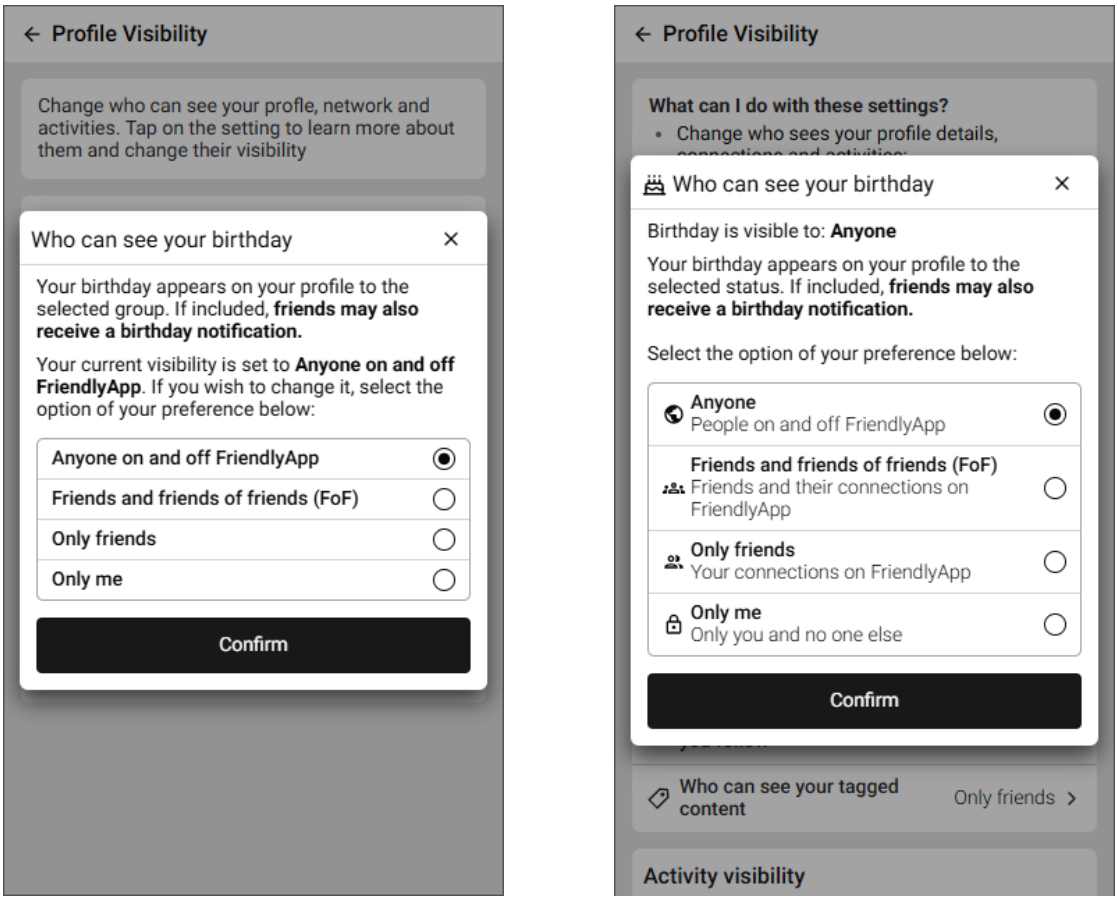
Table 23. "Expert A" evaluation on the Likert-scale questions.

	Q1. Clarity	Q2. Applicability	Q3. Validity
DP1	5.0	3.0	-
DP1AC1	-	-	2.0
DP2	5.0	3.0	-
DP2AC1	-	-	3.0
DP3	5.0	4.0	-
DP3AC1	-	-	2.0
DP3AC2	-	-	4.0
DP4	4.0	4.0	-
DP4AC1	-	-	3.0
DP5	2.0	2.0	-
DP5AC1	-	-	3.0
DP6	5.0	4.0	-
DP6AC1	-	-	4.0
DP6AC2	-	-	4.0
DP7	4.0	3.0	-
DP7AC1	-	-	3.0
DP8	5.0	4.0	-
DP8AC1	-	-	3.0
DP9	3.0	2.0	-
DP9AC1	-	-	2.0
DP10	5.0	3.0	-
DP10AC1	-	-	3.0
DP11	5.0	3.0	-
DP11AC1	-	-	3.0

Table 24. "Expert B" evaluation on the Likert-scale questions.

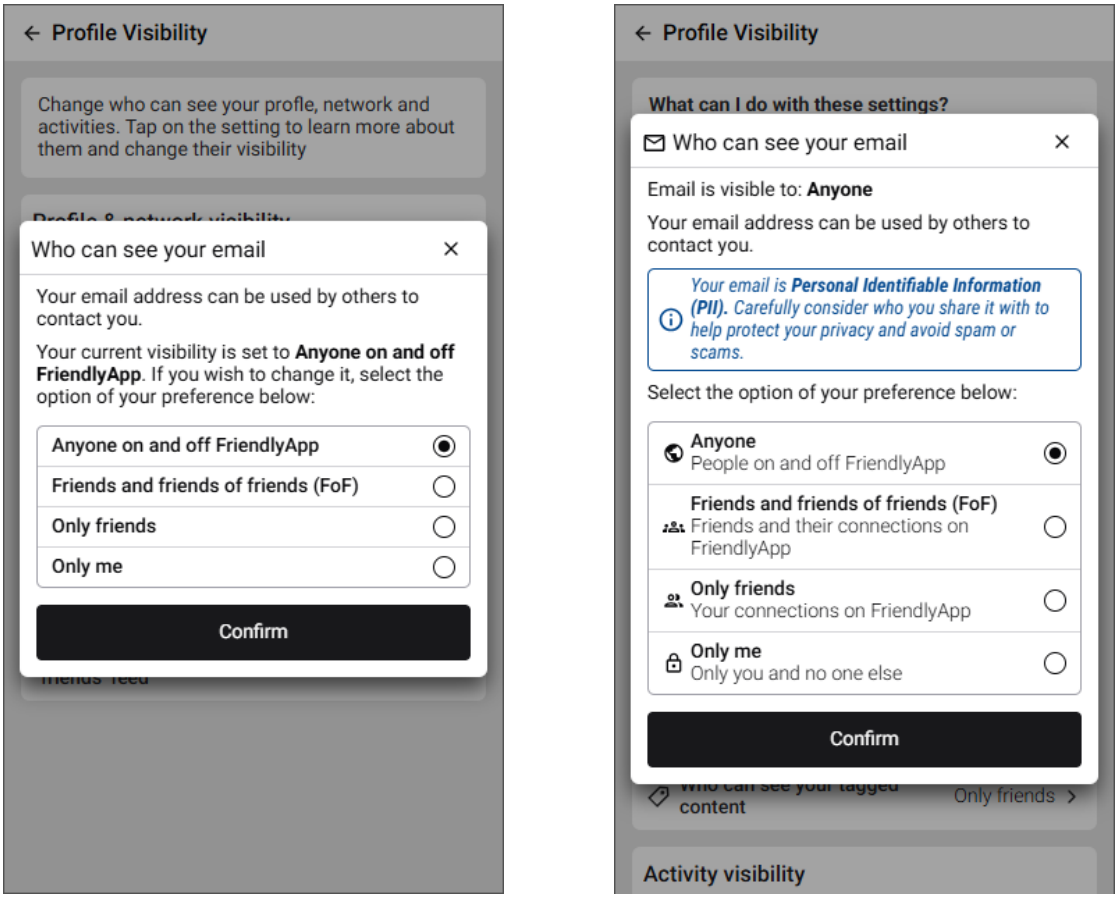
	Q1. Clarity	Q2. Applicability	Q3. Validity
DP1	5.0	4.0	-
DP1AC1	-	-	4.0
DP2	5.0	4.0	-
DP2AC1	-	-	4.0
DP3	5.0	5.0	-
DP3AC1	-	-	4.0
DP3AC2	-	-	4.0
DP4	5.0	4.0	-
DP4AC1	-	-	3.0
DP5	5.0	5.0	-
DP5AC1	-	-	4.0
DP6	4.0	4.0	-
DP6AC1	-	-	4.0
DP6AC2	-	-	4.0
DP7	5.0	5.0	-
DP7AC1	-	-	4.0
DP8	4.0	4.0	-
DP8AC1	-	-	5.0
DP9	4.0	3.0	-
DP9AC1	-	-	4.0
DP10	5.0	5.0	-
DP10AC1	-	-	5.0
DP11	5.0	4.0	-
DP11AC1	-	-	4.0

C Profile visibility settings remaining interfaces



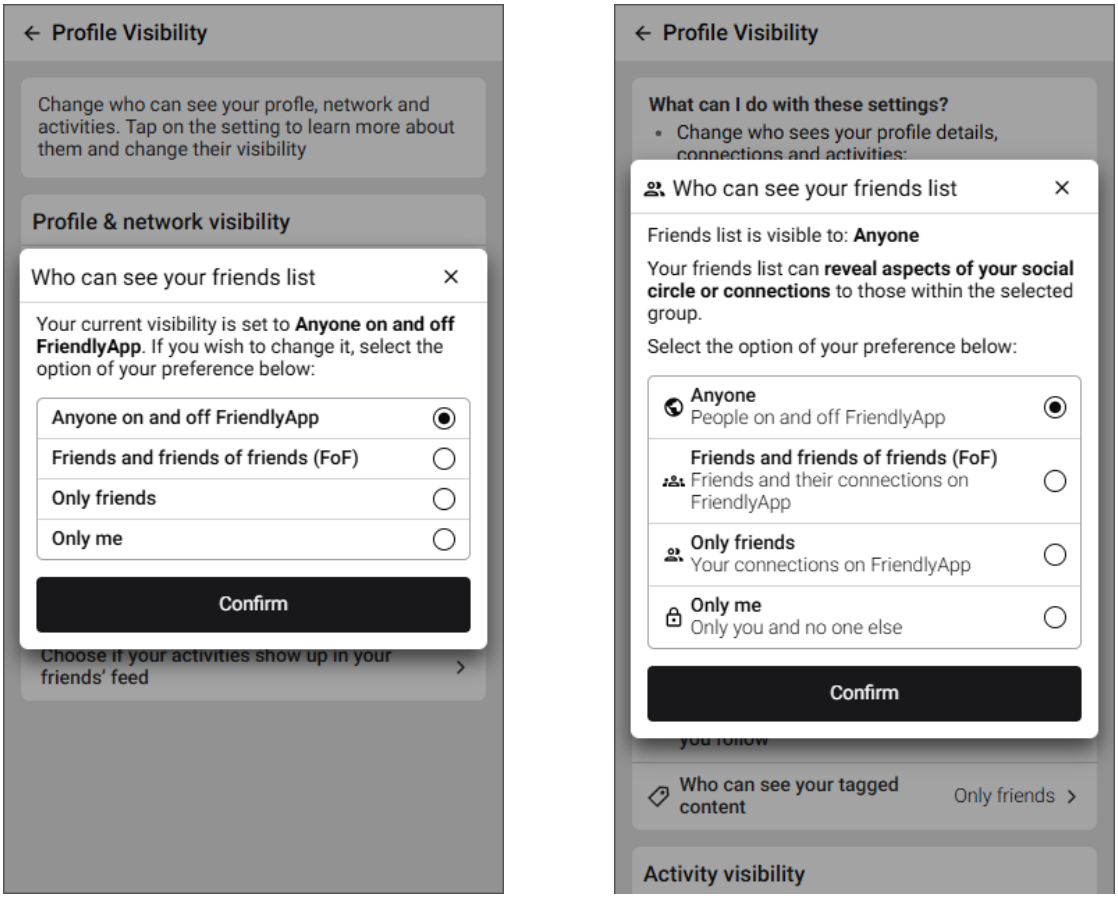
(a) (b)

Figure 23. *Who can see your birthday* (a) PH version and (b) RPH version.



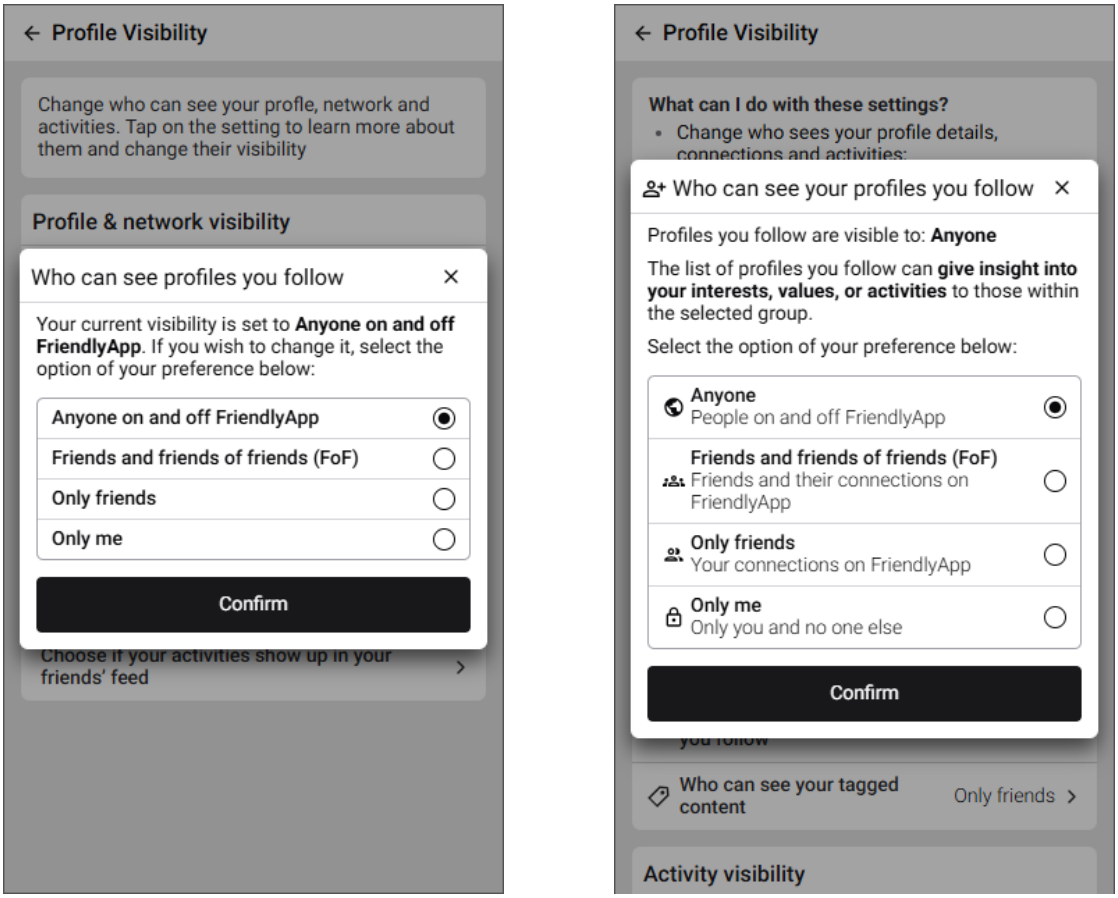
(a) (b)

Figure 24. *Who can see your email* (a) PH version and (b) RPH version.



(a) (b)

Figure 25. *Who can see your friends list* (a) PH version and (b) RPH version.



(a) (b)

Figure 26. *Who can see profiles you follow* (a) PH version and (b) RPH version.

D End-user experiment Google Forms

Privacy Usability Design Experiment

This experiment is part of a Master's thesis titled "An Approach for Designing Responsible Privacy Heuristics". The experiment focuses on evaluating different versions of a privacy setting. The researcher will be observing the session to understand how the participant interacts with the interface of a privacy setting. The participant will read a scenario, change the setting according to the scenario and respond to a questionnaire. Please keep in mind that our main aim is to evaluate the design, not the participant.

The participant's activities are described below:

1. Read a scenario and let the researcher know when you are ready to start the experiment.
2. You will receive access to an application to complete the scenario-based task. The scenario text will remain visible in a split-screen view for reference.
3. Perform the task at your own pace. Let the researcher know when you finish.
4. Answer a short post-experiment questionnaire.

* Indicates required question

Data collection

We are committed to respecting your privacy. We will not collect any Personal Information such as your name, address, or email, and all information we collect will not be linked back to you. Here's what we will collect, how, and why:

- **Demographic and background information**
  - **What:** age, gender, occupation, education level, familiarity with social media;
  - **How:** the researcher will ask you and note this information in this document.
  - **Why:** to understand whether factors like age, gender and etc, might influence how people interact with the settings.
- **Screen recording**
  - **What:** a video of your screen while you perform the tasks (no audio);
  - **How:** we will use screen recording software during the task phase.
  - **Why:** to help us analyze how participants interact with the interface and notice details we might otherwise miss;
- **Questionnaire answers**
  - **What:** your answers to the questionnaires (the researcher may ask clarifying questions if needed);
  - **How:** you will answer directly in the document provided.
  - **Why:** to help us evaluate the designs and understand your experience.

1. **Do you think using social media may negatively affect your privacy?** *

Mark only one oval.

- ◯ Yes, I believe it affects my privacy
- ◯ No, I do not believe it affects my privacy

Demographic and Background Information

2. **Age** * Dropdown

Mark only one oval.

- ◯ 18-25
- ◯ 26-35
- ◯ 36-45
- ◯ over 45

3. **Gender** *

Mark only one oval.

- ◯ Woman
- ◯ Man
- ◯ Non-binary
- ◯ Other: _____

4. **Occupation** *

Mark only one oval.

- ◯ Full-time employed
- ◯ Part-time employed
- ◯ Student
- ◯ Homemaker
- ◯ Retired
- ◯ Unemployed

5. **Education level** * Dropdown

Mark only one oval.

- ◯ No formal education
- ◯ Primary education
- ◯ Secondary education / High school
- ◯ Bachelor's degree
- ◯ Master's degree
- ◯ Doctoral degree (PhD, EdD, etc.)

6. **How long have you been using social media?** *

Mark only one oval.

- ◯ I've never used social media
- ◯ Less than 2 years
- ◯ Between 2 to 5 years
- ◯ Between 5 to 10 years
- ◯ More than 10 years

Scenario: Reviewing Your Profile Visibility Settings

Imagine you've recently signed up for a new social media platform called *FriendlyApp*. This app allows you to create a profile, share and react to posts, connect with friends, tag others, and follow profiles that match your interests and activities.

So far, you've mostly been using FriendlyApp to stay in touch with friends. However, you're now planning to use your profile to connect with professional contacts as well.

Since creating your account, you haven't reviewed your privacy settings. With this new context in mind, you want to **make sure that the information visible** on your profile **reflects the image you're comfortable sharing**.

Please take a moment to **review and adjust** (if required) your **profile visibility settings** according to your current preferences.
You will see the following settings available for configuration:

- Who can see your profile description
- Who can see your birthday
- Who can see your email
- Who can see your friends list
- Who can see profiles you follow
- Who can see your tagged content
- Who can see your posts
- Choose if your activities show up in your friends' feed

Once you've finished reading this scenario, let the moderator know so the experiment can begin.

Questionnaire

7. **Choose the version as instructed by the moderator.** *

Mark only one oval.

- ◯ Version A
- ◯ Version B

8. **How easy or difficult was it to use this interface?** *

Mark only one oval.

- 1 - Very difficult
- 2 - Slightly difficult
- 3 - Neutral
- 4 - Somewhat easy
- 5 - Very easy

9. **Were there any parts of the interface or information that made the decision difficult or confusing?** *

Please describe.

10. **How confident are you that the option you selected reflects your actual privacy preference?** *

Mark only one oval.

- 1 - Not confident at all
- 2 - Slightly unsure
- 3 - Neutral
- 4 - Somewhat confident
- 5 - Confident

11. **Did you feel that the privacy settings interface offered enough information for you to confidently decide who should see your profile details?** *

If not, what was missing?

12. **To what extent did you feel pressured by the information in the privacy setting interface to choose or not choose a specific option?**

Mark only one oval.

- 1 – Very strongly pressured (I felt I had little to no freedom to choose anything else)
- 2 – Strongly pressured (The information or interface clearly pushed me toward / against a specific option)
- 3 – Moderately pressured (I felt some pressure that may have influenced my choice)
- 4 – Slightly pressured (I noticed a small amount of influence, but it didn't affect my decision)
- 5 – Not at all pressured (I felt completely free to make any choice without influence)

13. **To what extent did the information provided in the privacy settings interface help you understand the potential consequences of your choices (e.g., privacy risks)?**

Mark only one oval.

- 1 – Not at all (I didn't feel more informed about potential consequences)
- 2 – Slightly (I got a little bit of insight)
- 3 – Somewhat (I understood some of the possible consequences)
- 4 – Mostly (The information gave me a good understanding)
- 5 – Completely (I clearly understood the potential consequences of each choice)

I. Acknowledgments

I would like to acknowledge the use of OpenAI ChatGPT[7] to help improve the clarity and formatting of the text in certain sections of this thesis.

II. Licence

Non-exclusive licence to reproduce thesis and make thesis public

I, **Beatriz Pontes da Costa Reis**,
*(*author's name)

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to

   reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

   An Approach for Designing Responsible Privacy Heuristics,
   *(*title of thesis)

   supervised by Mohamad Gharib.
   *(*supervisor's name)

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Beatriz Pontes da Costa Reis
11/08/2025

[7] https://chatgpt.com/