Baltic J. Modern Computing, Vol. 14 (2026), No. 2, pp. 462–491
https://doi.org/10.22364/bjmc.2026.14.2.10

# An Approach for Designing Responsible Privacy Heuristics

Beatriz PONTES DA COSTA REIS, Mohamad GHARIB

University of Tartu, Estonia

{beatriz.pontes.da.costa.reis,mohamad.gharib}@ut.ee

ORCID 0009-0001-1208-2715, ORCID 0000-0003-2286-2819

**Abstract.** Privacy compliance is a critical requirement for legal entities handling personal data (PD), demanding the integration of protective mechanisms into business workflows and transparency with data subjects (DSs). However, a critical gap persists: DSs struggle to understand privacy information and effectively use available protections. This human-centered challenge demands alignment between organizational processes and individuals' cognitive and behavioral capacities. Privacy heuristics (PHs) can bridge this gap by supporting user decision-making, yet their design is complex, prone to bias, and, if done irresponsibly, may lead to unethical or manipulative outcomes. This paper presents design principles for creating Responsible Privacy Heuristics (RPHs) in usable privacy-aware systems. We demonstrate applicability through online social network examples and validate through an A/B test with 12 users. Results show RPHs matched standard PHs in usability while proving slightly more effective at preventing privacy-invasive choices and fostering informed decision-making, without compromising user autonomy.

**Keywords:** Usable Privacy, Responsible privacy heuristic, Responsible design, Privacy Engineering, Privacy-aware systems

## 1 Introduction

In 2009, European Commissioner Meglena Kuneva famously stated that *"Personal data is the new oil of the Internet and the new currency of the digital world"* (Kuneva, 2009). This metaphor has only grown more pertinent, as scholars like Spiekermann et al. (Spiekermann et al., 2015) have highlighted the rise of complex personal data (PD) (also called Personal Information (PI)) markets. They, along with others (Gharib, 2022a), emphasize that PD has become a critical asset, powering services from personalized advertising and recommendation systems to risk analysis. In many contexts, particularly on social networks, the data itself is the product, generated by users and monetized by platforms.

The transformation of PD into a core product and service underscores the critical imperative to embed privacy protections within digital systems (Pattakou et al., 2018). In response, a global regulatory landscape has emerged, codified in laws and regulations such as the European Union's General Data Protection Regulation (GDPR) (European Parliament, 2016), Brazil's General Personal Data Protection Law (LGPD) (D'Oliveira and Cunha, 2024), and Japan's Act on the Protection of Personal Information (APPI) (Iwase, 2019), among others. These frameworks establish legal obligations for organizations to safeguard data subjects (DSs) by preventing various forms of PD mismanagement, including its misuse, excessive processing, improper storage, and unauthorized third-party sharing (Gharib et al., 2021).

Although legal entities handling PD must provide DSs with privacy protection mechanisms and disclose PD processing, the burden of understanding this information and using the mechanisms falls upon DSs (Jacobs and McDaniel, 2022). This creates a significant challenge, as users possess varying levels of digital literacy and often struggle with the legal jargon common in privacy policies and interface cues. This challenge reflects a broader issue: the design of processes involving PD must account for social and human factors. Traditional business processes prioritize efficiency and compliance, but privacy-aware systems must also address the psychological and behavioral dimensions of user interaction (Gharib, 2022b, 2025). Users are not passive endpoints; they are active participants with diverse expectations, preferences, and vulnerabilities. Ignoring this reality can result in user disengagement, eroded trust, and tangible harm.

A promising solution lies in the use of privacy heuristics, which can help users make informed decisions and take appropriate actions (Gharib, 2024). However, the design of these heuristics is inherently complex and prone to bias (Hjeij and Vilks, 2023). More critically, if not designed responsibly, they can unethically manipulate user judgment, leading to decisions that are immoral or socially irresponsible.

This paper aims to mitigate the burden Data Subjects (DSs) face when interacting with privacy mechanisms. To achieve this, we develop an approach centered on design principles for creating and evaluating Responsible Privacy Heuristics (RPHs). These principles guide designers in crafting privacy-aware solutions that empower users, uphold autonomy, and facilitate informed decision-making. This work extends our previous research (Da Costa Reis and Gharib, 2025) by: (1) refining our initial design principles and developing acceptance criteria (AC) for each principle; (2) validating both through expert review; (3) demonstrating the approach on realistic online social network examples; and (4) empirically validating it via A/B testing. Unlike prior work that focuses on identifying dark patterns or usability heuristics in isolation, this work integrates privacy heuristics, ethical design principles, and Design Science Research into a unified framework for designing RPHs.

The rest of this paper is structured as follows: Section 2 outlines the research baseline covering key concepts related to privacy heuristics and an analysis of both ethical and unethical design patterns. Section 3 outlines the methodology used to develop the approach, followed by a detailed description of the approach design in Section 4. Section 5 demonstrates the applicability of this approach, while Section 6 discusses the validation process. Section 7 lists and discusses the threats to the validity of this study, and finally, Section 8 concludes the paper and discusses future work.

## 2 Baseline

### 2.1 Heuristics & Privacy Heuristics

Heuristics are cognitive "rules of thumb" or "mental shortcuts" that facilitate faster decision-making (Hertwig and Pachur, 2015; Hjeij and Vilks, 2023). While they do not always guarantee optimality (Hjeij and Vilks, 2023), they are effective problem-solving tools for both well-defined and ill-defined problems by significantly reducing cognitive effort (Gharib, 2024). These shortcuts can be either instinctive (automatic) or deliberate, with experience enabling a transition between the two over time (Hjeij and Vilks, 2023).

Heuristics play a key role in online privacy, particularly in PD disclosure, where they are termed Privacy Heuristics (PHs). Sundar et al. (Sundar et al., 2020) categorize them by context: Personal, Social, and Technological. Complementing this, Marmion et al. (Marmion et al., 2017) propose six classes (e.g., Prominence, Network, Reliability) describing the cognitive principles underlying these decisions. In summary, privacy heuristics simplify complex decisions via mental shortcuts shaped by cues, experience, and biases. Table 1 synthesizes key heuristics from (Sundar et al., 2020; Marmion et al., 2017; Kitkowska, 2023) that influence privacy decision-making.

**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. |
| **Instant gratification.** People prioritize quick rewards at the expense of future gains. |
| **Loss aversion.** People prefer avoiding losses rather than acquiring 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 vouching for security influences disclosure. |

### 2.2 Unethical & ethical design patterns

To understand responsible privacy heuristics, we first examine ethical dimensions of decision-support mechanisms, analyzing both manipulative and ethical design patterns. We begin with unethical approaches before presenting responsible alternatives.

**Unethical patterns** (dark patterns) were coined by Brignull (Mathur and Mayer, 2021) as "tricks that make you do things you didn't intend to." These patterns are coercive and manipulative, guiding users toward decisions that benefit service providers at the user's expense (Caragay et al., 2024). They exploit biases and heuristics by hiding privacy-preserving options and promoting greater PD disclosure (Potel-Saville and Da Rocha, 2024). Kitkowska (Kitkowska, 2023) organized these patterns into taxonomies. Table 2 presents examples of privacy-deceptive patterns (PDPs), the heuristics they trigger, and their impact on users. While deceptive patterns are well-researched, a notable gap remains in research on what should be done (Caragay et al., 2024).

Table 2: Privacy Deceptive Patterns, heuristics and effects on user (Kitkowska, 2023)

| PDP | Heuristic(s) | Effect on user |
|---|---|---|
| **Privacy Zuckering:** is the use of deceptive design or persuasive tactics to manipulate users into sharing more PD than they intend to. | Choice overload, Status quo, Framing. | Users share more PD than they intended and might be unaware of how their PD is being processed. |
| **Bad defaults:** user account options are often preconfigured in a privacy-invasive manner (over-sharing) and sometimes with no alternative options available. | Default effect, Status quo, Loss aversion | 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 to use a service. | Instant gratification | Users end up sharing more PD than they intended to keep using a service. |
| **Trick questions:** deceive users into making privacy-invading choices with misleading or ambiguous wording. | Framing and anchoring. | Users will be confused and will most likely misinterpret their choices. |
| **Attention diversion:** distracts users from privacy-conscious choices by other aspects of the interface. | Anchoring and Framing | Hinders users from properly reflecting on their privacy-related choices. |
| **Confirmshaming:** steer users to make specific choices through guilt/shame. May use UI elements to induce a certain emotional state. | Affect, contrast and default effects | Users are manipulated to share more data through guilt or social pressure. |
| **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 PD than they intended to enable their accounts. |

**Ethical patterns,** also known as fair, or responsible patterns, are decision support mechanisms designed to prioritize the user's interests over businesses, enabling them to make informed and unobstructed decisions (Potel-Saville and Da Rocha, 2024). They function as the direct antithesis to dark patterns, which are characterized by user manipulation. The main objective of ethical patterns is user empowerment, achieved through

the transparent and clear presentation of choices. Their design must therefore adhere to principles of succinctness, transparency, and accessibility. A key contribution in this area is the taxonomy by Potel-Saville and Rocha (Potel-Saville and Da Rocha, 2024) (see Table 3), which maps dark patterns (DPs) to corresponding fair patterns (FPs).

**Table 3.** A taxonomy of dark and corresponding fair patterns that can influence privacy decisions

| Dark pattern | Fair pattern |
|---|---|
| **Harmful Default:** default settings are against the user's interest. | **Protective Default:** Defaults prioritize user privacy, well-being, and societal good. |
| **Missing Information:** Selective disclosure of information. | **Adequate Information:** Clear, sufficient, and relevant information with no overload. |
| **Maze:** User path to information, preferences, or choices are made unnecessarily complex. | **Seamless Path:** The user's path to information or choices is equally straightforward whether serving their interest or the provider's. |
| **Push & Pressure:** Emotional or time-based triggers pressure user decisions. | **No Pressure:** No manipulative nudges unless they serve user or societal benefits. |
| **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 UI is designed to mislead or trap users. | **Fair UX:** The UI ensures the clarity, shape, size, and prominence of buttons and icons. |

## 3 Design Science Research Methodology for Developing Responsible Privacy Heuristics

The research approach was developed following the Design Science Research (DSR) methodology (Peffers et al., 2007), which aims to *"improve the state of practice and contribute to design knowledge through the systematic construction of useful artifacts"* (Smuts et al., 2022). DSR not only focuses on the creation of such artifacts but also emphasizes their demonstration, evaluation, and communication. Our approach (illustrated in Fig. 1) aligns with DSR's key steps while further refining some into sub-steps, and described as follows:

1. **Problem identification:** This phase involves recognizing a gap in current practices, a deficiency in existing solutions, or an opportunity for improvement that can be addressed through the design of a novel artifact. The problem must be clearly defined and motivated to establish the necessity and relevance of the research. As discussed earlier, a clear gap exists between the prevalence of manipulative designs and the lack of prescriptive frameworks for ethical alternatives. Therefore, this research is motivated by the need for a structured approach to guide the design and evaluation of RPHs for privacy-aware solutions.
2. **Approach design:** This phase involves the creation of a purposeful artifact (e.g., a method, model) to address the identified problem. Our approach to designing the

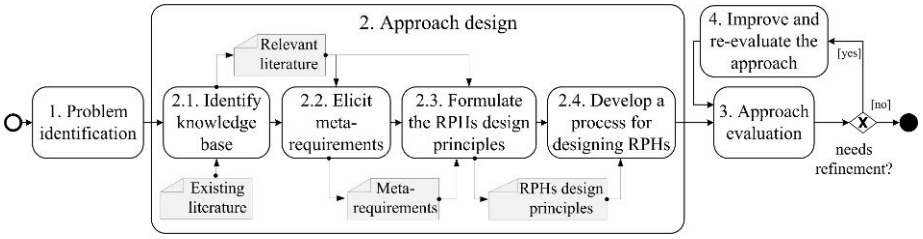



**Fig. 1.** The process for constructing and evaluating the approach

RPH framework 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 (Möller et al., 2020), and the last step offers a systematic process for using the approach. The approach design is further described in Section 4.

3. **Approach evaluation:** This phase is critical for validating the utility and efficacy of the proposed artifact. It involves gathering empirical evidence to assess how well the artifact—in this case, our RPH design approach—solves the problem identified in the first phase. This is achieved by defining specific, measurable evaluation criteria and employing appropriate methods to test the artifact against them. Our evaluation aims to assess the approach based on how well it supports the creation of solutions in the problem space, and it is further described in Section 6.
4. **Improve and re-evaluate the approach:** this step mainly focuses on identifying limitations or areas of improvement and refining the approach accordingly.

## 4 Approach design

This section details the methodological procedures for each step.

**1. Identify knowledge base.** The approach guides RPH design and evaluation for privacy-aware solutions. Design principles are prescriptive statements that aid requirement-to-design transfer and enable reuse across similar contexts (Cronholm and Göbel, 2018; Möller et al., 2020). In this step, we identified literature on ethical and unethical design patterns applicable to privacy heuristics, summarized earlier.

**2. Elicit Meta-requirements.** Drawing on literature, we define RPHs as user-empowering, transparent, accessible, and easy-to-understand decision-support mechanisms that respect user autonomy and enable informed privacy choices. They must be grounded in ethical principles (Respect, Beneficence, Non-maleficence (Kisselburgh and Beever, 2022), Justice, Integrity, Social Responsibility (Renaud and Shepherd, 2018)). Based on this, we elicited six Meta-Requirements (MR) listed in Table 4.

While these meta-requirements are intentionally formulated at a high level of abstraction, as they define the design objectives within the DSR process, we extend them with corresponding *Design Considerations* (see Table 4). These provide concrete guidance on how each abstract requirement can be translated into design decisions (e.g.,

avoiding biased defaults, ensuring reversibility of choices), thereby clarifying their role in guiding the derivation of actionable design principles.

**Table 4.** Meta-requirements for RPHs

| Meta Requirement (MR) - Source |
|---|
| **MR1. Integrity:** Information presented through RPHs must be accurate, truthful, consistent, and free from incomplete representations (Renaud and Shepherd, 2018; Kisselburgh and Beever, 2022). |
| *Design Considerations:* Ensure that information is not selectively omitted; claims (e.g., "data is not shared") remain consistent across the interface and are verifiable where applicable |
| **MR2. Non-manipulation:** A RPH must not exploit cognitive biases or limitations to steer users toward privacy-invasive decisions (Ahuja and Kumar, 2022). |
| *Design Considerations:* Avoid default options that bias toward privacy-invasive choices and eliminate dark patterns such as urgency cues or emotionally loaded language. |
| **MR3. Beneficence and non-maleficence:** A RPH should maximize user benefits while minimizing privacy risks and potential harms, ensuring ethical and responsible guidance (Renaud and Shepherd, 2018; Kisselburgh and Beever, 2022). |
| *Design Considerations:* Present privacy options with a clear balance between benefits and risks, avoid high-risk defaults, and clearly communicate potential harms. |
| **MR4. Autonomy and control:** A RPH should empower user's autonomy and freedom of choice by offering genuine, meaningful, and informed options without implicit coercion (Kisselburgh and Beever, 2022; Ahuja and Kumar, 2022). |
| *Design Considerations:* Provide equally accessible options and enable users to easily modify or revoke their choices without unnecessary barriers. |
| **MR5. Context-aware and accessible:** A RPH should, when possible, adapt to different contexts by considering risk levels, situational factors, and user diversity (Sundar et al., 2020). |
| *Design Considerations:* Provide additional explanations or warnings for sensitive data and ensure accessibility for users with varying levels of digital literacy and contexts of use. |
| **MR6. Regulatory compliant:** A RPH should align with applicable data protection regulations (European Parliament, 2016). |
| *Design Considerations:* Ensure that consent mechanisms are informed, specific, freely given, granular, and revocable in accordance with applicable regulations (e.g., GDPR). |

**3. Formulate the RPHs design principles.** The RPH design principles were formulated from the literature reviewed in the *Identify knowledge base* step. We analyzed deceptive patterns and their heuristics (Table 2) and prior research (Mathur and Mayer, 2021; Bösch et al., 2016; Gunawan et al., 2022; Ahuja and Kumar, 2022) to understand how user privacy behaviors are influenced. We drew particular insight from Ahuja and Kumar (Ahuja and Kumar, 2022), who identify 25 dark strategies and seven ethical concerns (e.g., compulsion, lack of control) grounded in four autonomy dimensions. Building on this, we formulated design principles aligned with our meta-requirements.

We then derived acceptance criteria (AC) as binary (Yes/No) checklists for each principle (Table 5), demonstrated in Section 5 and evaluated by experts (Section 6.1).

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

| |
|---|
| **DP1. Neutral:** A RPH should present information about privacy choices in a neutral, balanced, and clear 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 information:** 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. Instead, it should allow users to deliberate freely, ensuring informed and voluntary choices. |
| **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 PD; 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 prioritize user benefits while proactively minimizing 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)? |

**4. Developing a methodological process for designing RPHs.** The methodology for designing RPHs (Fig. 2) guides the design of RPHs, and comprises four key steps:

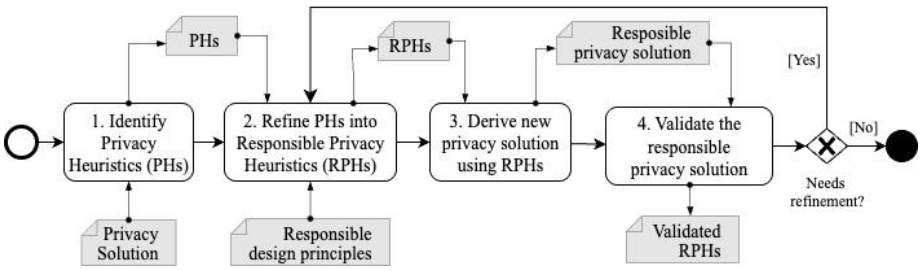



**Fig. 2.** The methodological process to be followed during the overall RPHs design

1. **Identify core privacy heuristics for usability:** This step takes the privacy solution as input and derives Privacy Heuristics (PH) from Gharib's ten Usable Privacy Heuristics (UPHs) (Gharib, 2024): Visibility, Revocability, Clarity, Expressiveness, Learnability, Minimalist design, Errors, Satisfaction, User suitability, and User assistance. These ensure privacy interactions are intuitive and user-friendly.
2. **Refine PHs into RPHs:** This step transforms PHs into RPHs using the responsible design principles. Designers select relevant principles and consult the acceptance criteria to guide refinement.
3. **Derive a new privacy solution based on RPHs:** Employs the selected RPHs and acceptance criteria to guide the revised design. Designers evaluate the solution using yes-or-no questions; positive responses satisfy the corresponding principles.
4. **Validate the responsible privacy solution:** Evaluates whether RPHs fulfill their purpose via end-user testing, expert reviews, or other assessment techniques, individually or in combination.

## 5 Applying the approach to an example from the online social network domain

This section demonstrates our approach by redesigning a social media platform's privacy settings—a user's primary tool for controlling how personal data is shared, accessed, and processed. To establish a realistic scope, we analyzed the core privacy functionalities of major Online Social Networks (OSNs), including Facebook, Instagram, LinkedIn, and Reddit. This review identified recurring features and control patterns, which informed the design of our example platform. Generally, OSNs allow users to: create personalized profiles; share content via posts; and interact through reactions, comments, and direct messaging. Table 6 outlines the privacy settings that correspond to these core functionalities in our example.

**Table 6.** 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); |

To avoid redundancy among overlapping settings, we focus on the first category from Table 6: *Profile Visibility*. Following our approach, we designed five distinct privacy solutions (available in Appendix A) for this category, each with a unique interface. Their descriptions are listed in Table 7. This paper details the application of our approach to the *Profile Visibility Default Interface*; the full set of applications is available online[1]

**Profile visibility default interface.** In what follows, we go through the methodological process for the *Profile visibility default interface* privacy solution.

In **Step 1**, we use the default interface of the *Profile Visibility* settings as input, illustrated in Fig. 3a, then we identify the core privacy heuristics that enhance usability. We identified minimalist interventions that provides DSs with essential information for

[1] https://hdl.handle.net/10062/117140

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 Fig. 3a), the DS is presented with a brief description of what these settings enable. DSs are also informed that tapping on a setting will lead them to more information about it—a feature 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*. Following this description, the privacy settings are split into two sections with descriptive names that avoid technical terms. This design choice accounts for the fact 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*.

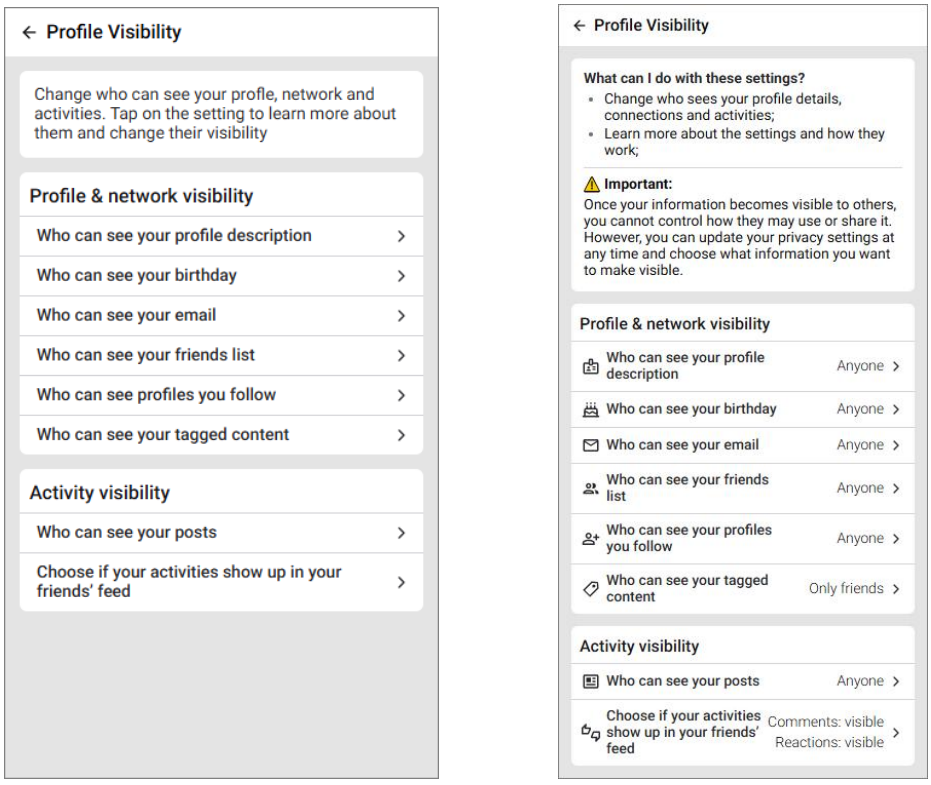
In **Step 2**, we refine the identified PHs into RPHs by applying relevant design principles. This process is documented in Table 8. For each PH, we first show its original form, followed by its refined version. Changes introduced by applying each principle are highlighted in **bold** (e.g., RPH1.1[2]).

With the PHs now refined into RPHs, we proceed to **Step 3** by revisiting the initial design solution. The updated interface, which incorporates these refinements, is presented in Fig. 3b. The following analysis examines the new design, detailing the implemented changes and the specific RPHs that informed them. Specifically, this revision transforms the description into a question to engage DSs. The content is presented in two bullet points to clearly highlight the enabled actions, adhering to **RPH6.1**'s principle of minimal, navigable design. A warning has been added to alert DSs to the privacy risks of making information visible. This aligns with **RPH1.2** by providing guidance, helping them reflect on potential consequences.

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

**Table 7.** 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. |




(a) Profile visibility following PH

(b) Profile visibility following RPHs

**Fig. 3.** Profile visibility settings following PHs and RPHs

After the description, the two-section structure is retained for consistency, but the settings descriptions have been enhanced with visual icons, and the current status is now immediately visible, removing the need for DSs to open each setting. This addition of at-a-glance visual cues is guided by **RPH6.1**, which enhances navigation through recognizable elements, and **RPH9.1**, which reinforces inclusive and accessible design. When implemented correctly, these elements support DSs in quickly locating and understanding the purpose of each control.

By displaying setting statuses directly on the *Profile Visibility* interface, we enable DSs to quickly review their current privacy choices and assess if changes are needed. This immediate visibility sets clear expectations about the available options and improves engagement, a change grounded in **RPH9.1**, **RPH6.1**, and **RPH1.2**. Furthermore, while the design presents more information, we have carefully balanced the layout to avoid overwhelming the DS or appearing to favor specific privacy actions, in strict adherence to **RPH4.1**'s principle of unbiased design.

The final step involves evaluating the design against the acceptance criteria of the applied principles. Table 9 lists how the AC for each principle has been achieved.

**Table 8.** 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.**" |

**Table 9.** Profile visibility default interface evaluation with AC.

| |
|---|
| **DP1AC1.** - Achieved: *Yes*. All links to the privacy settings follow the same pattern: icon, name, and current status. |
| **DP3AC1.** - Achieved: *Yes*. The privacy settings are one click away from the interface, and their links were updated with recognizable visual elements (e.g., icons) as well as the current setting status. |
| **DP3AC2.** - Achieved: *Yes*. Icons were chosen to resemble the functionality they are associated with (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.** - Achieved: *Yes*. 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.** - Achieved: *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.** - Achieved: *Yes*. The enhanced design maintains simple language, avoiding technical terms; utilizes different font weights to create contrast, and visual elements to make the design more engaging and sections of information more identifiable. |

# 6 Validation

This section validates the proposed approach in terms of its design principles and their AC through expert review, and its effectiveness for producing RPHs for usable privacy-aware systems via end-user experiment. Consistent with early-stage Design Science Research (DSR), this evaluation is exploratory in nature. It aims to provide initial evidence of the feasibility and potential usefulness of RPHs, rather than to establish statistically generalizable conclusions.

## 6.1 Validation of the design principles via experts

To validate the design principles, we engaged privacy experts to complete a structured questionnaire (Available online[3]). The questionnaire evaluated the **clarity** and **applicability** of each design principle, the **validity** of acceptance criteria, and the **completeness** of the proposed design principles and their AC. The questionnaire began with an introduction to the research context and objectives, followed by the evaluation methodology. Experts then assessed each principle and its AC, presented in a table, using a 5-point Likert scale (1=negative, 5=positive). The table included dedicated spaces for specific feedback, and the survey concluded with an open-ended question on completeness. For analysis, we defined three color-coded evaluation categories based on score ranges, summarized in Table 10, to determine satisfactory outcomes and facilitate an at-a-glance interpretation of the results.

**Table 10.** Evaluation categories for Likert-scale questions.

| Color | Score Range | Evaluation |
|---|---|---|
| Green | $\geq 4$ | Positive evaluation — the principle or criterion is considered clear, applicable, and valid. |
| Yellow | $\geq 3$ 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. |

Two privacy experts completed the questionnaire, and their evaluations are referred to as *Expert A* and *Expert B* for clarity. Table 11 presents the average scores from their questionnaire responses, color-coded according to the evaluation categories defined in Table 10. We interpret the quantitative scores alongside the experts' qualitative feedback where available. The experts evaluated the principles without a detailed demonstration of their practical application, which may have contributed to some lower scores.

**DP9. Situation-aware** received the lowest applicability score (below 3.0) for applicability (Q2). While *Expert A* gave a low score without justification, *Expert B* (score: 3.0) noted that distinguishing it from **DP8. Context-aware** required clarification. We addressed this by prioritizing *Expert B*'s feedback and providing concrete examples to

[3] https://zenodo.org/records/17508891

**Table 11.** 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 |

improve distinction. To address this concern, Table 12 provides explicit differentiation between DP8. and DP9.

Several principles received neutral scores. For **DP1. Neutral**, *Expert B* suggested replacing "displayed" with "designed" in **DP1AC1**, while *Expert A* noted potential cultural gaps in privacy awareness among practitioners. For **DP2. Honesty and Clarity**, *Expert B* recommended minor wording improvements, though we considered some suggestions redundant given **DP10. Accessible and inclusive**.

Other neutral-scoring elements included **DP3AC1**, **DP4AC1**, **DP5. Benefit-risk balance**, and **DP5AC1**. For **DP3AC1**, *Expert A*'s feedback about assessment complexity led us to refine the criterion from "easy to learn and intuitive to use" to "easy to navigate to". Neutral scores for **DP4AC1** likely reflect the inherent subjectivity in evaluating coercive patterns. For **DP5** and **DP5AC1**, we addressed *Expert B*'s suggestion for clarification through practical demonstrations rather than just formal definitions. We additionally observed that DP5. and DP6. are conceptually close and may confuse practitioners. Both address risk communication but differ in timing and function. Consequently, we clarified these distinctions in Table 12 along with the DP8./DP9. differentiation.

**DP7AC1**, **DP9**, **DP9AC1**, **DP11. Regulation compliant**, and **DP11AC1** also received neutral scores primarily from *Expert A*, who did not provide any feedback to justify these evaluations. *Expert B* rated these positively and suggested evaluating **DP11** alongside other principles to prevent superficial compliance. The feedback provided for **DP9AC1** was already discussed when addressing the lowest score category. We consider the positively evaluated principles satisfactory and will not discuss them further.

**Table 12.** Differentiation of closely related design principles

| Principle | Focus | Example Application | Boundary |
|---|---|---|---|
| **DP8. Context-aware** | Static characteristics of the interaction environment (data sensitivity, regulatory requirements, platform type) | Displaying a stronger privacy warning when users share health data (high sensitivity) vs. basic profile information (low sensitivity) | Focuses on *what* is being shared and *where* |
| **DP9. Situation-aware** | Dynamic, real-time user state and interaction flow (user's current task, decision fatigue indicators) | Simplified options after 5+ consecutive decisions; pause confirmation after rapid, uncharacteristic changes | Focuses on *when* and *how* the user is interacting |
| **DP5. Benefit-risk balance** | Trade-off communication between positive outcomes and negative outcomes of a privacy choice | Sharing your location enables nearby friend alerts (benefit) but allows tracking of your movements (risk) | Concerns *content* of what is communicated |
| **DP6. Consequences awareness** | Post-decision feedback and long-term implication transparency | After selecting a setting: Your birthday is now visible to 'Everyone'. This information may be used for identity verification or targeted advertising. | Concerns *timing* (during/after) and *mechanism* of communication |

Experts also evaluated the principles' collective completeness and adaptability through open-ended questions. When asked, *"Do the current design principles, collectively, cover necessary aspects of responsible privacy? Are they adaptable across different privacy contexts...?"*, *Expert B* affirmed good coverage and suggested emphasizing integration throughout the Privacy by Design cycle. *Expert A* did not directly address completeness but noted in response to the open-ended prompt that regulatory compliance alone does not ensure ethical practice—a concern already reflected in **DP11**.

### 6.2 Validation of the responsible privacy solution via end-users

We validated the effectiveness of the approach through an A/B test with end-users, comparing a baseline interface (designed with existing PHs) against an enhanced version (designed with our proposed RPHs derived from the original PHs using our methodology). The test used the previously developed Profile Visibility settings, augmented with additional settings from the same group. Related interfaces were designed using Figma[4], a digital tool for design and prototyping. The design was then

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

implemented using NextJS, ChakraUI, and Typescript and deployed using Vercel[5]. In what follows, we present the validation objectives, outline the experimental method, and analyze the results.

**Objectives and Criteria.** This experiment evaluates whether the RPH-based design fosters user autonomy and enables more informed decisions without compromising usability. Consequently, we assess the following criteria: **1. Informed decision:** Compares the user's actual choice against an optimal or reference action. **2. Decision awareness:** Captures evidence of user interaction with informative elements (e.g., hovers, clicks, dwell time). **3. Perceived usability:** Measures how easily participants could use the interface, ensuring the RPH version performs at least as well as the PH baseline. **4. Perceived informed decision:** Assesses whether users felt they had sufficient information to make an informed privacy choice. **5. Perceived autonomy:** Measures the extent to which participants felt free from interface pressure or manipulation. **6. Perceived consequence-awareness:** Evaluates whether the interface helped users understand potential risks and outcomes. Please note that RPH-based design is not intended to restrict disclosure but to support alignment between user intentions and outcomes. Users may deliberately choose to share PD to obtain benefits; therefore, the goal is to ensure that such decisions are informed and deliberate rather than unintentionally influenced by interface design.

**Methodology.** We recruited 14 participants with diverse backgrounds and varying familiarity with social media privacy settings. Two participants were excluded: one due to a language barrier affecting interface interaction, and another who failed the quality control question requiring acknowledgment that social media can negatively impact privacy. The experiments were moderated and conducted primarily remotely, with a few sessions in person. The researcher initiates screen recording (using OBS Studio[6] for remote sessions; Mac's built-in recorder for in-person). Recording begins after informing the participant which design version they will use. Participants proceed with the task, notifying the researcher upon completion, at which point the recording stops. Each participant interacted with only one design version. The protocol for all sessions was as follows:

1. **Briefing:** Participants access the experiment's
   The Google Form[7], which details the purpose, data collection (demographics, screen recording, post-experiment questionnaire), and procedure. They read instructions at their own pace, and we answer any questions. While the duration was mentioned during recruitment, we reiterate that there is no time pressure. Participants must acknowledge that social media can affect privacy (only "Yes" responses proceed) and provide informed consent before continuing.
2. **Demographic questionnaire:** Participants provide their age range, gender, occupation, education level, and familiarity with social media.
3. **Experiment:** Involves the following steps

[5] https://vercel.com/
[6] https://obsproject.com/
[7] `https://zenodo.org/records/17508891`

   (a) **Scenario:** Participants read a scenario and notify the researcher when ready to begin. The researcher then provides the application link and instructs them to split their screen between the application and the scenario.
   (b) **Access interface:** Participants access the assigned interface version (Version A for the PH baseline or Version B for the RPH-enhanced interface).
   (c) **Review and adjust privacy settings:** Participants are asked to review and select the privacy option that best matches their preference for all four different interfaces. Then, inform the researcher upon task completion.
4. **Post-experiment questionnaire:** Participants return to the Google Form to complete the post-experiment questionnaire, which includes questions related to assess perceived usability, informed decision, autonomy, and consequence awareness.
5. **Respond to any clarifying questions:** the researcher conducts a brief follow-up discussion to clarify specific choices made during the interaction or responses provided in the questionnaire.

**Results.** We collected data via Google Forms, screen recordings, and post-experiment interviews, which included demographic information, social media habits, and participants' rationale for their choices. To structure our findings, we first present participant demographics, then analyze the privacy settings they selected, discuss post-experiment questionnaire responses, and finally, provide a discussion on the experiment's findings.

**Participant Demographics.** Our sample consisted of 12 participants; their demographics are summarized in Table 13. We also categorized participants as **active** or **passive** social media users based on their self-described behavior. **Active users** regularly create and share content and interact with others' posts, while **passive users** primarily consume content with minimal interaction or private sharing. Four of the six participants using the PH version were **active** users, compared to only one in the RPH group. All **active** users were women, and all men were **passive** users. Notably, even active users reported conservative sharing habits. Participant assignment to design versions (PH vs. RPH) was balanced by age and education, but not by expertise or social media behavior, which were collected post-experiment. This resulted in the PH group having five women and one man, and the RPH group having two women and four men.

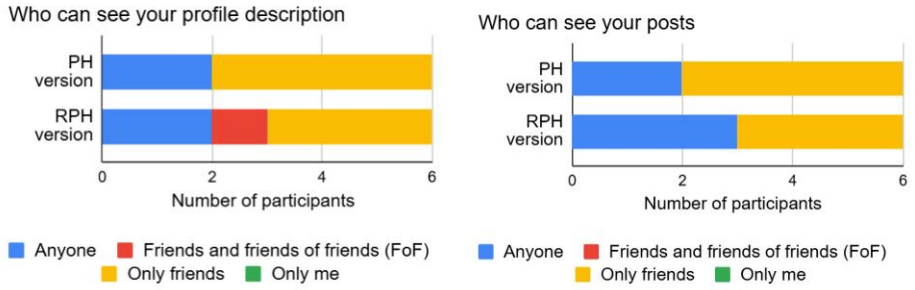
**Participants privacy choices.** The analysis of participant privacy choices and their interactions provides a multi-faceted comparison between the PH (baseline) and RPH (enhanced) interface versions. We focused on assessing **Informed decision** and **Decision awareness**, and participants were introduced to the following privacy configuration interfaces: 1. *Who can see your profile description?* (see 3a and 3b for PH-based and RPH-based design, respectively). 2. *Who can see your posts?* (see 10a and 10b for PH-based and RPH-based design, respectively). 3. *Who can see your birthday?* (see 11a and 11b for PH-based and RPH-based design, respectively). 4. *Who can see your email?* (see 12a and 12b for PH-based and RPH-based design, respectively). 5. *Who can see your friends' list?* (see 13a and 13b for PH-based and RPH-based design, respectively). 6. *Who can see profiles you follow?* (see 14a and 14b for PH-based and RPH-based design, respectively). 7. *Who can see your tagged content?* (see 15a and 15b for PH-

**Table 13.** Participants' background information.

| # | Age | Gender | Employment | Education | Social media familiarity |
|---|---|---|---|---|---|
| 1 | 26-35 | Man | Full-time | Bachelor's | > 10 years |
| 2 | 26-35 | Woman | Full-time | Bachelor's | > 10 years |
| 3 | 26-35 | Woman | Homemaker | Master's | > 10 years |
| 4 | 26-35 | Woman | Full-time | Master's | > 10 years |
| 5 | 18-25 | Woman | Student | High school | 5-10 years |
| 6 | 26-35 | Woman | Student | Master's | > 10 years |
| 7 | 26-35 | Woman | Part-time | Bachelor's | > 10 years |
| 8 | 26-35 | Man | Full-time | Bachelor's | > 10 years |
| 9 | 36-45 | Man | Full-time | Master's | > 10 years |
| 10 | 36-45 | Woman | Unemployed | Bachelor's | > 10 years |
| 11 | 18-25 | Man | Student | Bachelor's | 5-10 years |
| 12 | 26-35 | Man | Full-time | High school | > 10 years |

based and RPH-based design, respectively). 8. *Choose if your activities show up in your friends' feeds* (see 16 and 17 for PH-based and RPH-based design, respectively).

For profile and post visibility (Figures 4a and 4b), four participants (two per version) selected *Anyone* for professional reasons. While both groups made this choice, their processes differed: one RPH user deliberately switched to *Anyone* only after confirming optional fields, allowing sensitive data to be excluded. Screen recordings showed RPH users engaged in more deliberation, while a PH user decided quickly without reflection. This indicates the RPH version better promoted thoughtful **decision awareness**. Quantitatively, RPH users selected *Anyone* in 2/6 cases (33.3%) for profile visibility and 2/6 (33.3%) for post visibility, compared to PH users with 2/6 (33.3%) and 1/6 (16.7%), respectively.

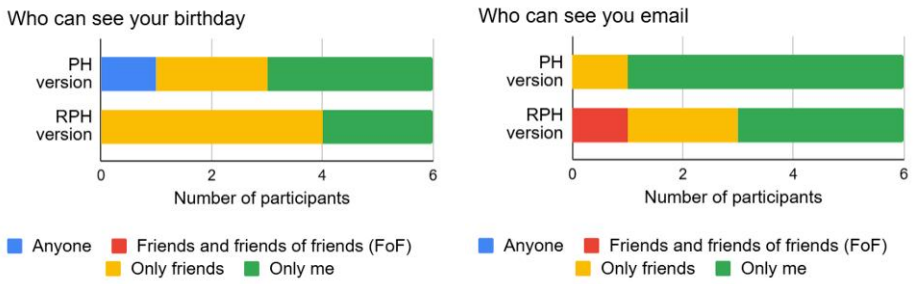



(a) (b)

**Fig. 4.** (a) who can see your profile description and (b) who can see your posts

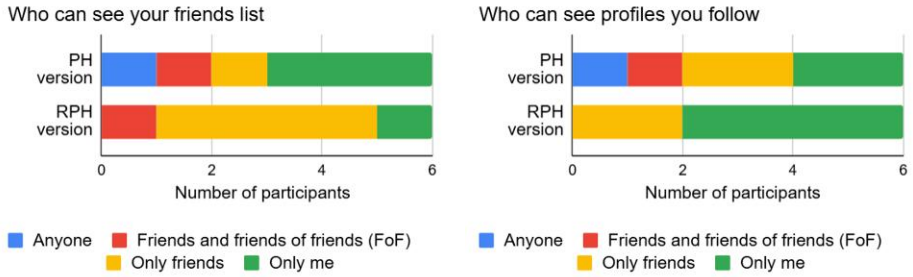
For birthday and email visibility (Figures 5a and 5b), the RPH version demonstrated superior support for informed choices. Notably, no RPH user selected the public **Any-**

**one** option for their birthday (0/6, 0%), unlike one PH user (1/6, 16.7%) who saw no risk. For email visibility, one RPH user (1/6, 16.7%) changed their selection from **Anyone** to a more private option after engaging with the interface; no PH user made such a change. This informed deliberation is reflected in the longer average time RPH users spent on the setting (M = 18.0s, SD = 6.2) compared to PH users (M = 8.0s, SD = 3.1), yielding a large effect size (Cohen's d = 2.06). For the friends list and followed profiles settings (Figures 6a and 6b), only a PH user selected the **Anyone** option (1/6, 16.7%), citing a professional profile strategy. The RPH version, despite minimal design differences from the PH version (see Figures 13 and 14), resulted in no users making this fully public choice (0/6, 0%).




(a) (b)

**Fig. 5.** Results from (a) who can see your birthday and (b) who can see your email.

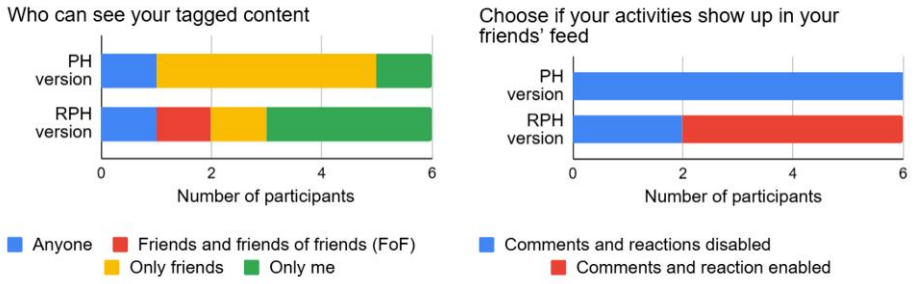



(a) (b)

**Fig. 6.** (a) who can see your friends list and (b) who can see profiles you follow.

For tagged content and activity feed visibility (Figures 7a and 7b), both versions had one user select **Anyone** for tagged content (1/6, 16.7% per version) based on misconceptions, indicating uninformed decisions. Despite the RPH user spending more time

on the settings (M = 22.4s, SD = 8.1) than the PH user (M = 11.2s, SD = 5.3), the risks were not effectively communicated. For the complex activity feed, the RPH version better maintained user engagement with clearer explanations. These results highlight a need for further refinement of the RPH design to enhance risk communication and decision awareness for these specific settings. Due to the exploratory, underpowered design (N=6 per condition), p-values and confidence intervals are not reported; presented statistics reflect directional trends only.

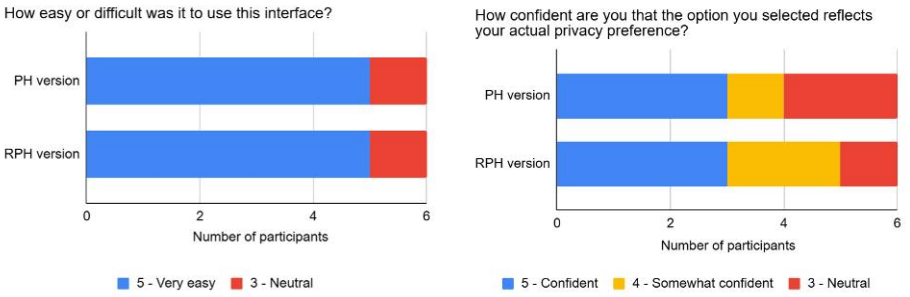



(a) (b)

**Fig. 7.** (a) who can see your tagged content and (b) choose if your activities show up in your friends' feed.

**Post-experiment questionnaire.** The questionnaire assessed perceived design aspects. To assess **perceived usability**, participants rated ease of use on a 5-point scale. The RPH version performed equally well as the PH version (Fig. 8a), despite containing more information. However, open-ended feedback revealed that some participants who rated usability as "very easy" still described both interfaces as text-heavy, suggesting information density affects perceived clarity independently of overall ease-of-use ratings. Specific critiques included unclear activity visibility (PH version) and uncertainty about tag functionality (RPH version). Concerning **Perceived informed decision**, measured on a 5-point confidence scale (1=Not confident at all, 5=Confident), the RPH version scored slightly higher than the PH version (Fig. 8b). While all participants agreed the interfaces provided sufficient information, neutral ratings were attributed to external factors: one PH user cited general platform distrust, while an RPH user reported infrequent social media use. One RPH participant reiterated concerns about tag functionality, suggesting a need for clearer descriptions.

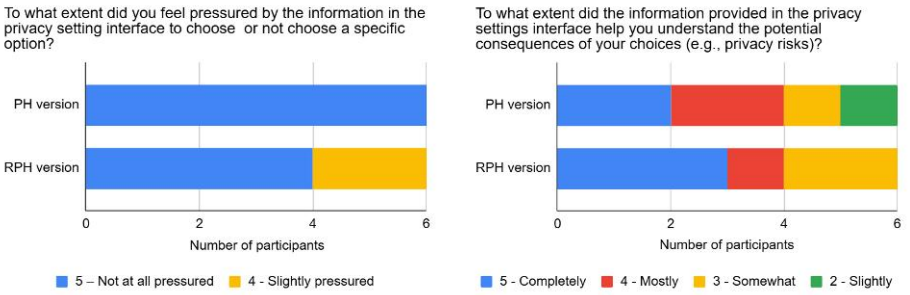
For **Perceived autonomy**, rated on a 5-point pressure scale (1=Very strongly pressured, 5=Not at all pressured), RPH users reported feeling slightly more influenced than PH users (Fig. 9a). This aligns with the RPH's design goal to encourage deliberation, as demonstrated by one user who changed their email setting from **Anyone** to **Friends and friends of friends (FoF)** after engaging with the interface guidance. Concerning **Perceived consequence awareness**, measured on a 5-point scale (1=Not at all, 5=Completely), the RPH version scored higher than the PH version (Fig. 9b). While pre-existing knowledge influenced scores, the RPH design's "soft warnings"




(a) Perceived usability results (b) Perceived informed decision results.

**Fig. 8.** Perceived usability and informed decision

proved effective: three participants specifically noted the informational cues prompted reflection, with one adopting a more conservative setting directly as a result.




(a) Perceived autonomy results (b) Perceived consequence awareness results.

**Fig. 9.** Perceived autonomy and consequence awareness

**Discussion of results.** Overall, the RPH version demonstrated several key advantages. It reduced selections of the most public option (**Anyone**) by 25% compared to the PH version. While both interfaces were equally **Perceived usable**, RPH significantly outperformed PH in fostering **Informed decision**, **Perceived informed decision**, and **Decision awareness**. Users of the RPH version reported higher **Perceived consequence awareness** and showed more deliberate engagement, with one user concretely changing from **Anyone** to **Friends and friends of friends** after deliberation. The increased influence reported in **Perceived autonomy** for RPH users reflects this positive, deliberate engagement rather than coercive pressure.

## 7 Threats to validity

This research contributes to the development of responsible privacy design, though several potential threats to validity should be acknowledged. Following Wohlin et al. (Wohlin et al., 2012), we classify threats to validity under:

1. **Internal validity** concerns whether the observed effects can be causally attributed to the experimental conditions. A key threat is the potential for participants to alter their behavior when aware of being observed. To mitigate this, we instructed participants to act naturally and assured them they were not being personally evaluated. A second threat is moderator bias, which was minimized by strictly limiting the moderator's role to answering procedural questions, with no guidance provided on interface usage or privacy choices.
2. **External validity** concerns the generalizability of our findings. The approach was evaluated using relatively simple privacy mechanisms in a generic social media context. More complex scenarios, such as cookie consent or privacy policy interfaces, remain unexamined. Future work should assess the approach's applicability across diverse privacy contexts and user groups, and validate its effectiveness in mitigating deceptive patterns within real-world applications. Moreover, the small sample size (n=12) limits statistical power and increases the risk of Type II errors. As a result, the study is not intended to establish statistically significant effects but rather to identify trends and generate hypotheses for future research.
3. **Conclusion validity** concerns the reasonableness of the study's conclusions. Given the exploratory nature of the validation and limited participant diversity, the findings regarding the effectiveness and usability of the responsible privacy heuristics indicate a modest positive impact. These findings should be interpreted as exploratory and preliminary. More extensive studies with larger, more diverse samples would strengthen confidence in these conclusions.

## 8 Conclusions and future work

This paper presented an approach for designing Responsible Privacy Heuristics (RPHs) to support the creation of ethical and usable privacy mechanisms. Developed through design science research, the contribution comprises 11 design principles with acceptance criteria and a methodological process for application. This provides a practical framework for translating ethical requirements into privacy solution design. Validation via an A/B test demonstrated that the RPH-based design maintained usability while slightly enhancing privacy risk awareness and supporting more informed decision-making, without compromising user autonomy.

Future research should expand on the demonstration of the proposed approach by applying 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, and potentially find gaps and ways to improve the approach. To broaden the range of the proposed approach, 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. Finally, we will investigate the interdependencies between RPHs and system requirements (Gharib and Mirzazada, 2025). Since privacy decisions often affect usability, accessibility, transparency, security, and compliance, understanding these dependencies could help identify conflicts and trade-offs early and support more systematic and human-centered privacy-aware design.

## Acknowledgment

This work was supported by the Estonian Research Council grant "Developing human-centric digital solutions" (No. TEM-TA120).

## References

Ahuja, S., Kumar, J. (2022). Conceptualizations of user autonomy within the normative evaluation of dark patterns, Ethics and Information Technology **24**(4), 52.
https://link.springer.com/article/10.1007/s10676-022-09672-9

Bösch, C., Erb, B., Kargl, F., Kopp, H., Pfattheicher, S. (2016). Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns, Proceedings on Privacy Enhancing Technologies **2016**(4), 237–254.
http://c2.com/cgi/wiki

Caragay, E., Xiong, K., Zong, J., Jackson, D. (2024). Beyond Dark Patterns: A Concept-Based Framework for Ethical Software Design, Conference on Human Factors in Computing Systems - Proceedings, ACM, p. 16.
https://dl.acm.org/doi/abs/10.1145/3613904.3642781

Cronholm, S., Göbel, H. (2018). Guidelines supporting the formulation of design principles, ACIS 2018 - 29th Australasian Conference on Information Systems, Vol. 1.
https://www.diva-portal.org/smash/record.jsf?pid=diva2:1269110

Da Costa Reis, B. P., Gharib, M. (2025). Towards an Approach for Designing Responsible Privacy Heuristics, Joint Proceedings of RCIS (Research Challenges in Information Science) Workshops and Research Projects Track, Vol. 3987, pp. 1–15.
https://ceur-ws.org/Vol-3987/paper3.pdf

D'Oliveira, N., Cunha, F. J. A. P. (2024). Brazilian General Data Protection Law (LGPD): the relationship between information policy and information regime, Revista Digital de Biblioteconomia e Ciencia da Informacao **22**.
https://www.scielo.br/j/rdbci/a/DWntpkXMB9GgCPKycFcxtts/?lang=en

Parliament, E. (2016). 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 Da, Official Journal of the European Communities **59**, 1–88.

Gharib, M. (2022a). Privacy and Informational Self-determination Through Informed Consent: The Way Forward, Lecture Notes in Computer Science, Vol. 13106 LNCS, Springer Science and Business Media Deutschland GmbH, pp. 171–184.
https://link.springer.com/chapter/10.1007/978-3-030-95484-0_11

Gharib, M. (2022b). Toward an architecture to improve privacy and informational self-determination through informed consent, Information and Computer Security **30**(4), 549–561.

Gharib, M. (2024). Towards a Heuristic Model for Usable Privacy, Joint Proceedings of RCIS (Research Challenges in Information Science) Workshops and Research Projects Track, CEUR-WS.org, pp. 1–10.
https://ceur-ws.org/Vol-3674/ASPIRING-paper3.pdf

Gharib, M. (2025). From Perception to Protection: A Mental Model-Based Framework for Capturing Usable Security and Privacy Requirements, The 30th Nordic Conference on Secure IT Systems (NordSec25), pp. 465–483.
https://link.springer.com/chapter/10.1007/978-3-032-14782-0_25

Gharib, M., Giorgini, P., Mylopoulos, J. (2021). COPri v.2 — A core ontology for privacy requirements, Data and Knowledge Engineering **133**, 101888.
https://linkinghub.elsevier.com/retrieve/pii/S0169023X2100015X

Gharib, M., Mirzazada, E. (2025). ReInTa: A Novel Requirements Interdependencies Taxonomy, Baltic Journal of Modern Computing **13**(4), 834–861.
https://www.bjmc.lu.lv/fileadmin/user_upload/lu_portal/projekti/bjmc/Contents/13_4_05_Gharib.pdf

Gunawan, J., Santos, C., Kamara, I. (2022). Redress for Dark Patterns Privacy Harms? A Case Study on Consent Interactions, Proceedings of the 2022 Symposium on Computer Science and Law, Association for Computing Machinery, Inc, pp. 181–194.
https://dl.acm.org/doi/abs/10.1145/3511265.3550448

Hertwig, R., Pachur, T. (2015). Heuristics, History of, International Encyclopedia of the Social and Behavioral Sciences: Second Edition, pp. 829–835.
https://pure.mpg.de/rest/items/item_2139307/component/file_2404607/content

Hjeij, M., Vilks, A. (2023). A brief history of heuristics: how did research on heuristics evolve?, Humanities and Social Sciences Communications **10**(1).
https://www.nature.com/articles/s41599-023-01542-z

Iwase, H. (2019). Overview of the act on the protection of personal information, European Data Protection Law Review **5**(1), 92–98.

Jacobs, D., McDaniel, T. (2022). A Survey of User Experience in Usable Security and Privacy Research, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 13333 LNCS, Springer Science and Business Media Deutschland GmbH, pp. 154–172.
https://link.springer.com/chapter/10.1007/978-3-031-05563-8_11

Kisselburgh, L., Beever, J. (2022). The ethics of privacy in research and design: Principles, practices, and potential, Modern Socio-Technical Perspectives on Privacy, pp. 395–426.
https://library.oapen.org/bitstream/handle/20.500.12657/52825/1/978-3-030-82786-1.pdf#page=392

Kitkowska, A. (2023). The Hows and Whys of Dark Patterns: Categorizations and Privacy, Human Factors in Privacy Research pp. 173–198.
https://library.oapen.org/bitstream/handle/20.500.12657/76226/978-3-031-28643-8.pdf?sequence=1#page=177

Kuneva, M. (2009). Roundtable on online data collection, targeting and profiling.

Marmion, V., Bishop, F., Millard, D. E., Stevenage, S. V. (2017). The cognitive heuristics behind disclosure, decisions, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 10539 LNCS, Springer Verlag, pp. 591–607.

Mathur, A., Mayer, J. (2021). What makes a dark pattern... dark? design attributes, normative considerations, and measurement methods, Conference on Human Factors in Computing Systems - Proceedings, Association for Computing Machinery, pp. 1–18.

Möller, F., Guggenberger, T. M., Otto, B. (2020). Towards a Method for Design Principle Development in Information Systems, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 12388 LNCS, Springer Science and Business Media Deutschland GmbH, pp. 208–220. https://link.springer.com/chapter/10.1007/978-3-030-64823-7_20

Pattakou, A., Mavroeidi, A. G., Kalloniatis, C., Diamantopoulou, V., Gritzalis, S. (2018). Towards the design of usable privacy by design methodologies, Proceedings International Workshop on Evolving Security and Privacy Requirements Engineering, ESPRE, pp. 1–8. https://ieeexplore.ieee.org/abstract/document/8501325/

Peffers, K., Tuunanen, T., Rothenberger, M. A., Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research, Journal of Management Information Systems **24**(3), 45–77. http://www.tandfonline.com/doi/full/10.2753/MIS0742-1222240302

Potel-Saville, M., Da Rocha, M. (2024). From Dark Patterns to Fair Patterns? Usable Taxonomy to Contribute Solving the Issue with Countermeasures, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 13888 LNCS, Springer Science and Business Media Deutschland GmbH, pp. 145–165.

Renaud, K., Shepherd, L. A. (2018). How to make privacy policies both GDPR-compliant and usable, International Conference on Cyber Situational Awareness, Data Analytics and Assessment, CyberSA, pp. 1–8. https://ieeexplore.ieee.org/abstract/document/8551442/

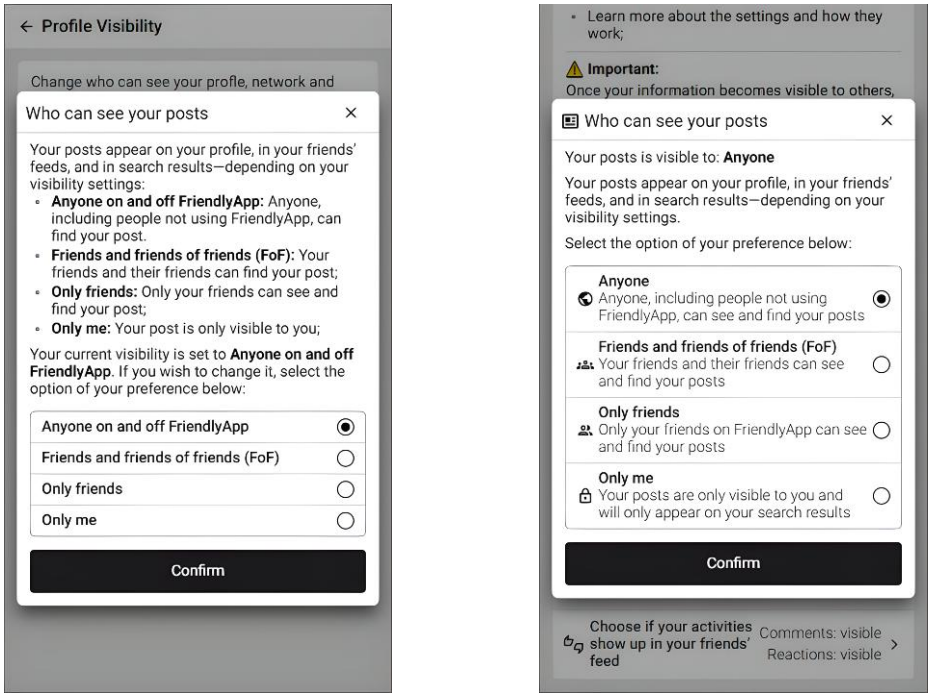
Smuts, H., Winter, R., Gerber, A., van der Merwe, A. (2022). "Designing" Design Science Research – A Taxonomy for Supporting Study Design Decisions, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 13229 LNCS, Springer Science and Business Media Deutschland GmbH, pp. 483–495. https://link.springer.com/chapter/10.1007/978-3-031-06516-3_36

Spiekermann, S., Acquisti, A., Böhme, R., Hui, K. L. (2015). The challenges of personal data markets and privacy, Electronic Markets **25**(2), 161–167. https://link.springer.com/article/10.1007/S12525-015-0191-0

Sundar, S. S., Kim, J., Rosson, M. B., Molina, M. D. (2020). Online Privacy Heuristics that Predict Information Disclosure, Conference on Human Factors in Computing Systems - Proceedings, Association for Computing Machinery, pp. 1–12.

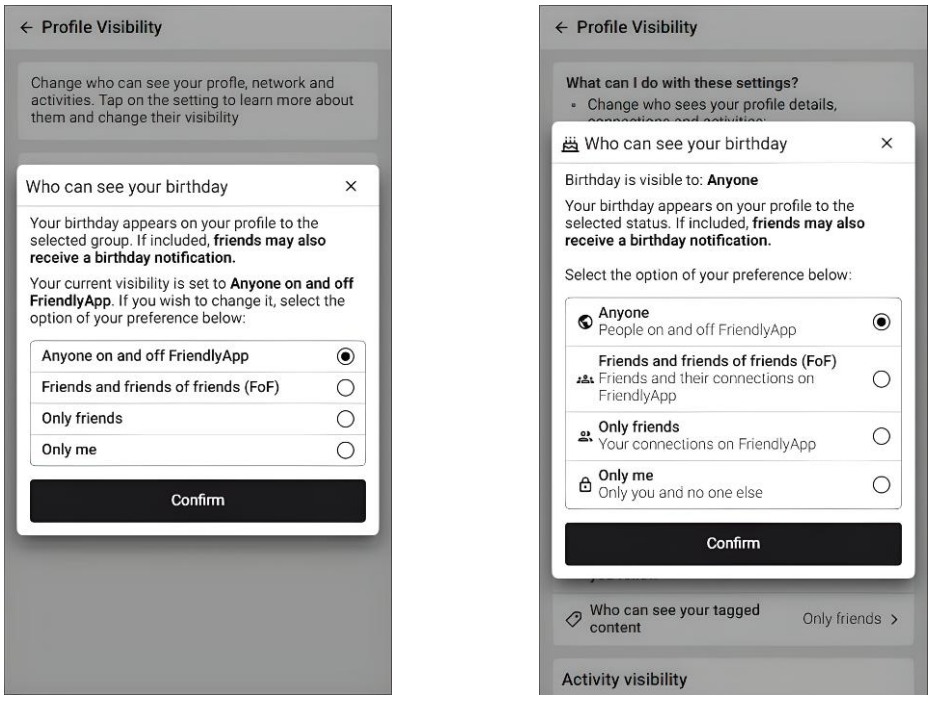
Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., Wesslén, A. (2012). Experimentation in software engineering, Vol. 9783642290, Springer Science and Business Media.

## Appendix A: Responsible privacy solutions used in demonstration and end-user validation



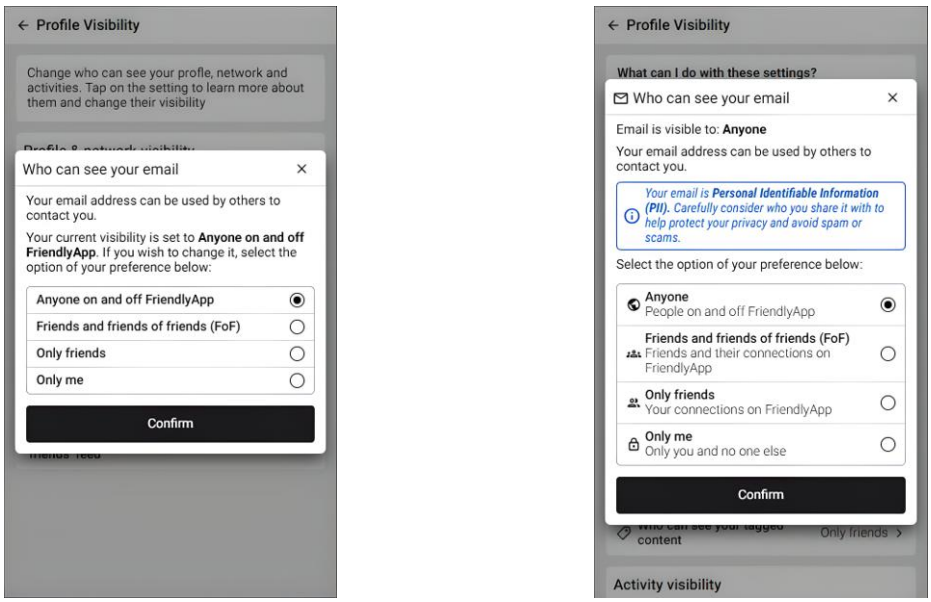

(a) (b)

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



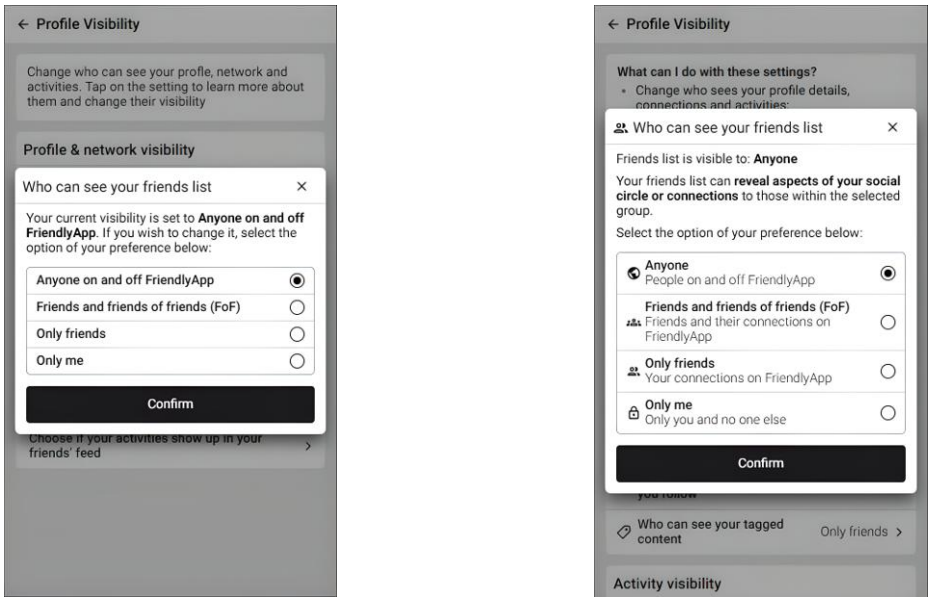

(a) (b)

**Fig. 11.** *Who can see your birthday* (a) PH version and (b) RPH version.



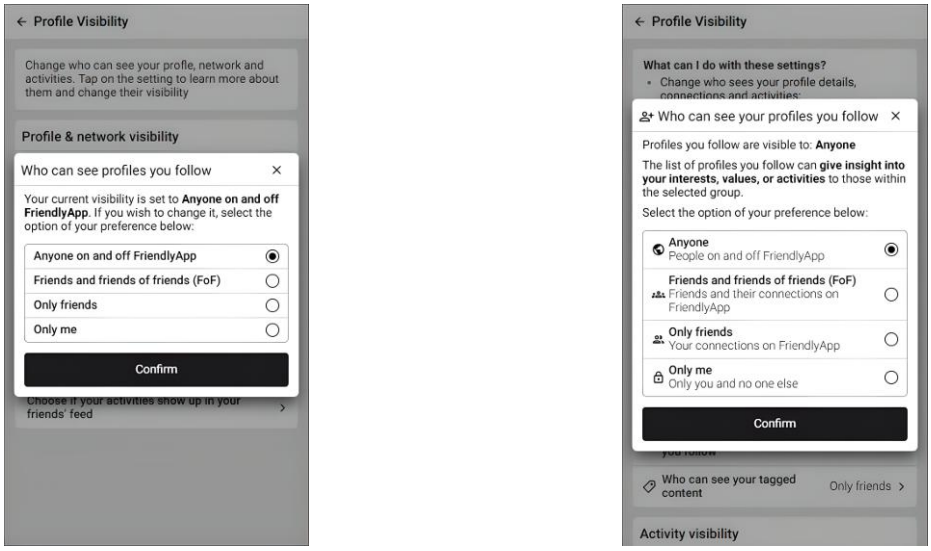

(a) (b)

**Fig. 12.** *Who can see your email* (a) PH version and (b) RPH version.



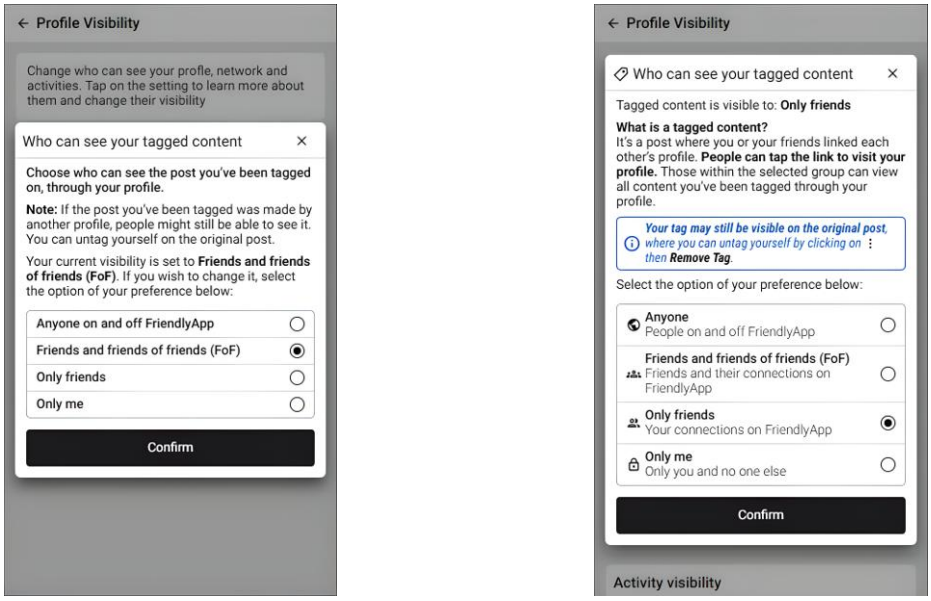

(a) (b)

**Fig. 13.** *Who can see your friends list* (a) PH version and (b) RPH version.



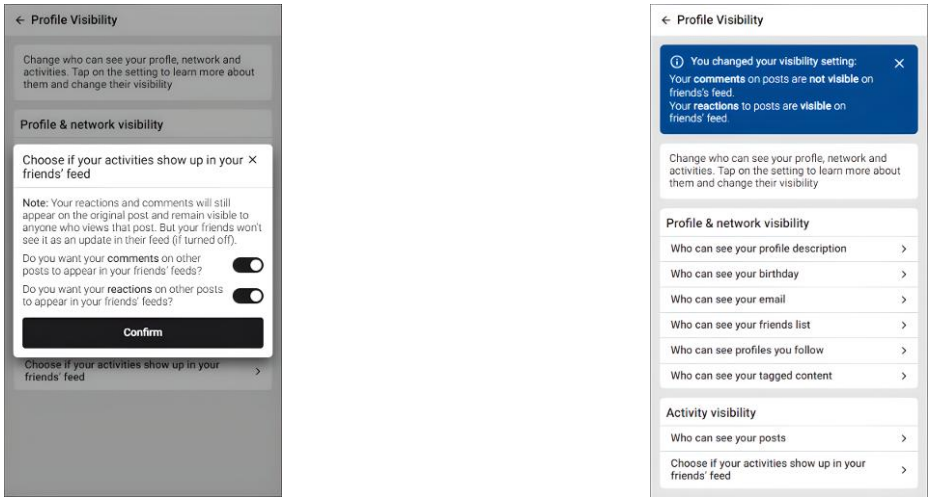

(a) (b)

**Fig. 14.** *Who can see profiles you follow* (a) PH version and (b) RPH version.



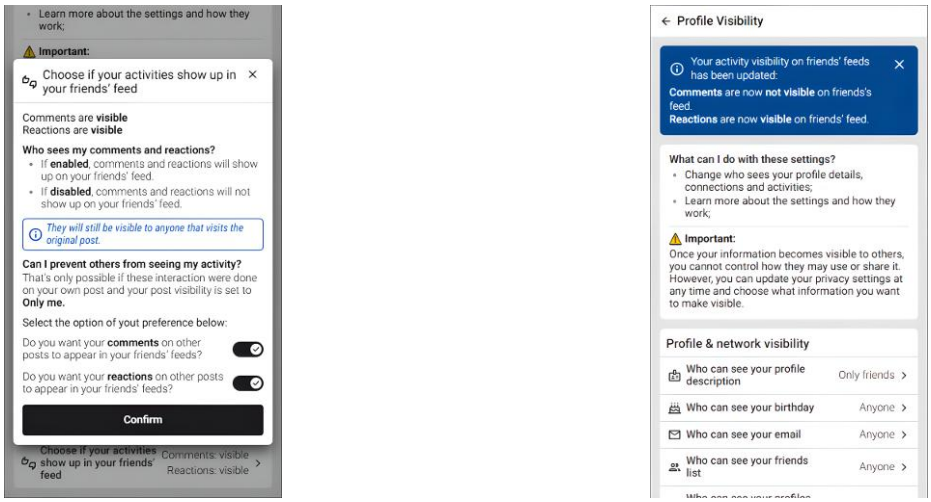

(a) (b)

**Fig. 15.** Who can see your tagged content (a) PH version and (b) RPH version.




(a) (b)

**Fig. 16.** *Choose if your activities appear in friends' feed* (a) setting and (b) feedback, RPH version.




(a) (b)

**Fig. 17.** *Choose if your activities appear in friends' feed* (a) setting and (b) feedback, RPH version.

Received February 19, 2026 , revised April 30, 2026, accepted June 18, 2026