

USENIX

THE ADVANCED COMPUTING
SYSTEMS ASSOCIATION

How Effective are Privacy Labels at Informing Users About App Data Handling Practices?

Sophia Walsh, *University of Bristol;* Lukasz Piwek, *University of Bath;*
Kopo Marvin Ramokapane, *University of Bristol*

https://www.usenix.org/conference/soups2026/presentation/walsh

This paper is included in the Proceedings of the
Twenty-Second Symposium on Usable Privacy and Security.

August 24–26, 2026 • Hannover, Germany

ISBN 978-1-939133-59-5

Open access to the Proceedings of the
Twenty-Second Symposium on Usable Privacy and Security
is sponsored by USENIX.

How Effective are Privacy Labels at Informing Users About App Data Handling Practices?

Sophia Walsh
University of Bristol

Lukasz Piwek
University of Bath

Kopo Marvin Ramokapane
University of Bristol

Abstract

Mobile app privacy labels aim to enhance transparency by helping users understand app data-handling practices. However, limited empirical evidence demonstrates whether they achieve this goal. Using a card-matching activity with think-aloud protocols, we examined how users interpret iOS Privacy Nutrition Labels and Android Data Safety Labels. Participants ($n = 16$) matched privacy labels to corresponding apps, revealing their existing knowledge of app data-handling practices and engagement patterns. Thematic analysis generated three themes: (1) Understanding – how users comprehend app data-handling practices; (2) Salience – which label elements participants notice; and (3) Impact – how awareness of actual data practices influences users. Our findings extend existing work on technical terminology comprehension, providing rich qualitative insights into user assumptions, heuristic reasoning, and emotional responses. While privacy labels can advance transparency goals, their current presentation undermines their utility. Future label designs should aim to dispel rather than compound misconceptions, develop heuristic abstractions to communicate information effectively, and reduce negative emotional impacts.

1 Introduction

Mobile phone applications (apps) play a pivotal role in modern society, enabling vital activities such as shopping, banking, security authentication, travel and communication. While data collection is essential for app functionality, it is also frequently collected for commercial reasons. With an estimated four million apps available to download [13], the sheer volume of data that can be collected and shared is vast. Given the diverse range of individual and societal behavioural insights and preferences that this data can reveal [10], its value [36], and the myriad of privacy risks [43], privacy and security are important considerations in the app selection process.

In 2020, Apple made it mandatory for all app developers to declare app-specific data-handling practices, presented in the form of 'privacy nutrition labels' (PNL). In 2022, Google adopted a similar approach, requiring the publication of 'data safety labels' (DSL). Both platforms claim these labels help users understand data practices and make informed decisions about app installation [1,19]. However, whether labels achieve these goals in practice remains an open question.

Research on privacy labels increases year on year [3], focusing on improving label form [49], proposing new applications for labels such as IoT [41], [9], and transparent AI [42]. Other research focuses on accountability - examining how accurate labels are [29, 30, 32, 40], identifying challenges for developers in creating accurate labels [27] and designing tools to support more accurate PNL [31, 47]. User studies find that people are largely unaware of privacy labels and struggle to understand the terminology used [25, 33, 48].

While research examines information recall regarding app data-handling practices, gaps remain in our understanding of how users actually apply this information. According to Bloom's Taxonomy, understanding is a hierarchy of cognitive skills starting with recall and ascending through comprehension, application, analysis, evaluation, and creation [37]. By framing privacy labels within this tiered structure, we can better evaluate how users comprehend and apply information on labels beyond recall and to what extent they can make informed decisions when choosing apps.

To address this gap, we ask the following research questions: (RQ1) what do users understand about privacy labels when asked to use them, (RQ2) how do users apply this knowledge, and (RQ3) what impact do privacy labels have on users' attitudes and intentions? To answer these RQs, we adopted an inductive, think-aloud method, in which 16 Android and iOS

Copyright is held by the author/owner. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee.
USENIX Symposium on Usable Privacy and Security (SOUPS) 2026.
August 23–26, 2026, Hannover, Germany.

users mapped privacy labels to real apps [45], enabling us to capture authentic user understanding and engagement while minimising demand effects [38]. Presenting participants with real data-handling practices allowed us to observe how users engage with privacy labels and the influence labels have on their decisions.

Using thematic analysis, we identified three themes: (1) Understanding - users hold incomplete knowledge of app data-handling practices so navigate labels using naive assumptions informed by reputation, media and experience; (2) Salience - participants rely on heuristic cues like data quantity when interpreting labels; and (3) Impact - exposure to actual data practices produces negative emotional and affective responses and motivates users to consider implications of data-handling. Overall, our findings suggest that in their current form, mobile app privacy labels are not effective in achieving their stated purpose - to inform user understanding of app data-handling practices. When people try to use them, they rely on heuristic reasoning and salient label features. These findings can help to inform future design iterations.

Key Contributions:

- Introduces a novel "applied-comprehension" methodology that evaluates how users actively use and interpret privacy labels, shifting the research paradigm beyond traditional metrics of recall.
- Demonstrates that users rely on heuristic shortcuts (such as list length and data-type volume) to evaluate privacy labels rather than reading them in full, even under optimal conditions.
- Provides qualitative evidence showing how pre-existing user assumptions about data-handling practices are shaped by external factors such as brand reputation, media exposure, and personal experience.
- Offers empirical insights into the psychological friction caused by accurate data transparency, showing that genuine comprehension of an app's data-handling practices negatively affects user attitudes and stated behavioural intentions.

2 Background and Related Works

2.1 The Usefulness of Privacy Labels

Privacy labels refer to a genre of privacy notice and choice mechanisms developed by privacy researchers [22]. Based on the notice-and-choice paradigm, they are mechanisms designed to communicate relevant privacy information to the user on a notice - to enable informed user choice [12]. Unlike regulatory consent dialogues or cookie banners, which must offer control over data collection and sharing [10], privacy labels are non-interactive information tools. They do not collect consent but rather aim to inform users about data practices and enable comparison between apps [1, 19].

Initially, privacy labels served dual purposes: to enhance the usability of privacy policies by highlighting salient information and reducing reading time by summarising key privacy facts, and to signpost users towards the full privacy policy [22]. In 2013, with increased app adoption, Kelley et al. [24] proposed their use specifically for mobile apps. Since 2020 and 2022 respectively, they have been adopted by Apple and Android for all apps uploaded on their platforms. Recently labels have also been proposed in other contexts. For instance, IoT devices [9, 14, 39, 41], smart pet technology [34], GDPR compliance [16], multidisciplinary communication spanning business, legal, regulatory, and human factors [20], and open-source generative AI applications [42].

2.2 The Usability and Effectiveness of Privacy Labels

In principle, privacy labels can advance user knowledge, encouraging consumers to consider data practices when selecting mobile apps and smart devices. Moreover, they may also increase privacy and security awareness more generally, empowering users to navigate the digital marketplace according to their individual preferences [21]. To fulfil this potential, privacy labels must be usable and effective at supporting real-world consumer decision-making.

The seminal work on privacy labels [22] evaluated their usability relative to privacy policies across three dimensions: information-finding accuracy, time, and enjoyability. Initial focus groups ($n = 24$) demonstrated that participants located and recalled data-handling information more accurately and quickly when using labels rather than policies while rating the experience was more enjoyable. These findings were later validated using a larger online sample ($n = 764$) [23] reinforcing the conclusion that labels are more usable than privacy policies. These studies established recall and information-finding as the dominant metrics for evaluating label usability, a precedent that has largely persisted in subsequent research.

Kelley et al. [24] first introduced the concept of mobile app privacy labels in 2013. Employing a two-part user study (laboratory $n = 20$ and online $n = 366$) to compare the impact of prototype labels (tabular privacy checklists) and the 2013 Android permissions display on app selection decisions, they found that the majority of participants did not notice or pay attention to privacy information. However, for those that did engage with the privacy information, labels had greater effect on app selection than the existing permissions display. This study also positioned privacy permissions alongside other factors influencing app selection: ratings, reviews, price, branding, word of mouth, number of downloads, popularity, app size, developer, and advertising. Privacy permissions ranked eighth out of 11 (above size and popularity), suggesting that even when labels are present, they compete with many other decision-making cues.

Following widespread platform adoption by Apple and

Google in 2020 and 2022, research attention increased. Cranor [12] identified that effective labels needed to be accurate, understandable, and actively used, calling for iterative design and evaluation. Zhang et al. [49] proposed an expandable grid prototype and found ($n = 200$) that it enabled users to answer key privacy questions more quickly and accurately than existing iOS labels, demonstrating that design improvements can meaningfully enhance usability performance. Research by Li et al. [33] compared the usability of PNL and DSL, assessing comprehension through definition tasks and fact-based questions. Neither was deemed to be more usable than the other but the need for increased clarity and comprehension for terminology, concise design, the inclusion of security information and the verification of labels was highlighted. Other research on the usability of PNL [48] proposed five criteria to evaluate usability and effectiveness: readability, comprehension, salience, relevance and control control. This criteria is adopted from research on the effectiveness of *privacy consent notices*. Using semi-structured interviews ($n = 24$) to explore user understanding of specific terms, usefulness, likeability and perceived future use, they found that users perceive labels to increase user comprehension but also report widespread misunderstandings and a lack of trust.

Balash et al. [2] showed that highly salient data collection disclosures such as linking financial data to user identities or tracking sensitive personal information significantly heighten a user's privacy and security risks and reduced their willingness to install the app. This suggests that direct engagement with label content can influence consumer choice, but raises the question of whether users typically engage deeply enough for this effect to occur in practice.

Despite widespread adoption, research consistently reports user misunderstanding of privacy labels [25, 33, 44, 48]. Most existing research evaluates understanding through recall-based measures. Our study extends this by using an applied-comprehension approach, exploring how users comprehend and apply label information when making app choices. By asking participants to use labels in a simulated choice activity, we observe how participants reason about app data-handling practices, offering insights into label usefulness. To our knowledge, only Zhang et al. [50] have examined whether labels address users' privacy concerns, finding that they only partially do so. Our study extends this by adopting an applied-comprehension approach examining what users understand when asked to use labels (RQ1), how users apply this knowledge (RQ2), and what impact privacy labels have on attitudes and intentions (RQ3).

3 Methodology

The study was conducted between March and September 2025. Using a think-aloud protocol [5], participants ($n = 16$) were audio-recorded while engaging in two tabletop matching activities. This section introduces our study design and materials, data analysis procedures, and limitations.

3.1 Ethics Statement

The study was reviewed and approved by our IRB. All participants provided informed consent, including explicit consent for audio recording. We used Microsoft-Teams to audio-record; no video was captured, and this was explicit to all participants. Microsoft-Teams was selected as our institutionally provisioned and supported recording tool, configured to meet GDPR and data protection requirements (i.e., encryption, access controls, managed storage). Before the study the Microsoft-Teams configuration was checked against these and IRB requirements. On completion of the study, the primary researcher manually transcribed all audio recordings and anonymised the data. Audio recordings were deleted after the coding was complete.

3.2 Study Design

The primary aim of this research was to explore what participants understand when they use privacy labels and whether privacy labels improve user understanding of app data-handling practices. However, simply asking participants what they understand about privacy labels and if engaging with them improves their understanding suffers from significant demand effects. We developed a methodological tool that enabled participants to engage with label information while mitigating these effects. Rather than asking participants to find or recall information, our methodology required our participants to match labels to apps. The study was scaffolded by the activity rather than through direct questions, facilitating free-flowing dialogue. Think-aloud protocols enabled us to observe how participants understood and used the information disclosed on the privacy labels.

This study design enabled us to observe user understanding of app labels beyond assessment of technical knowledge or recall ability. It also allowed us to observe which aspects of the labels stood out when users engage with them - that is, what users noticed and the most salient label features. In addition, it presented the opportunity for users to reflect on an app's actual disclosed data-handling practices. Study procedures and materials are described below.

3.2.1 Study Materials

We used a card-matching activity originally designed to facilitate conversations about data privacy and explore smartphone users' diverse experiences with app privacy [45]. The activity required participants to match apps with their corresponding privacy labels based on their understanding of app data-handling practices. By pairing the card-matching activity with a think-aloud protocol, this methodological approach

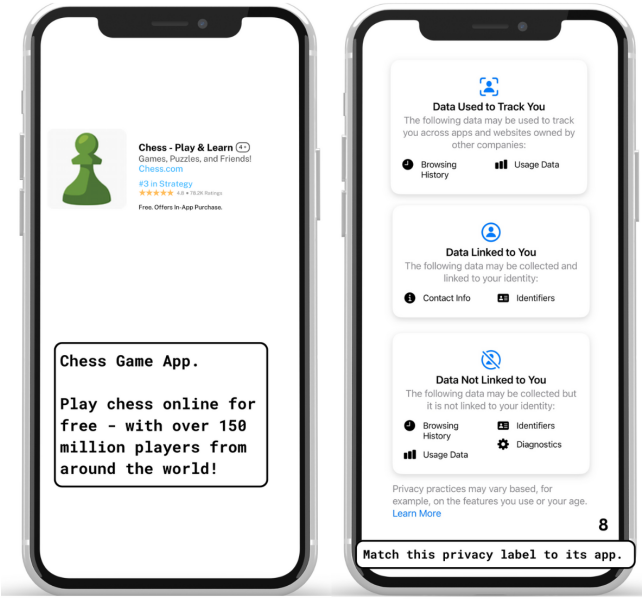


Figure 1: Example of iOS PNL and App Cards

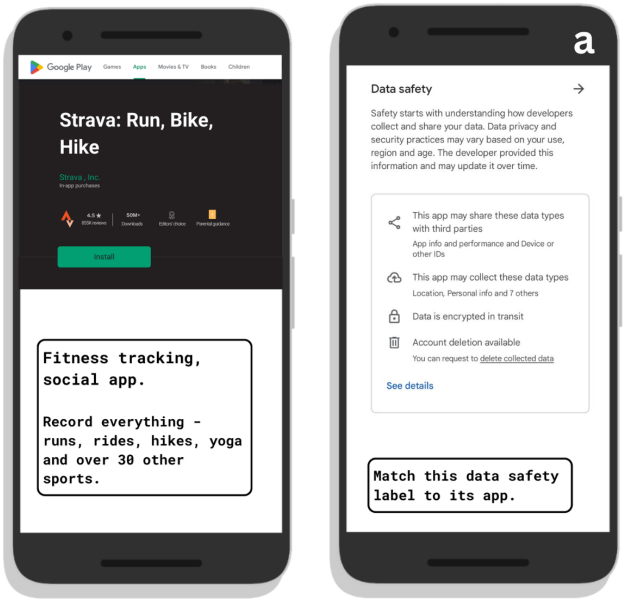


Figure 2: Example of Android DSL and App Cards

allowed us to capture genuine user comprehension, avoiding the need for direct questioning.

The card-matching tool comprised of 18 cards: nine cards depict smartphone apps (app cards) and nine featuring their corresponding privacy labels (label cards). There are two versions of the tool - one for Android and one for iOS - each featuring the same nine apps. Each app card replicates how the app appears in the platform-specific app store, including the app's purpose, genre, star rating, download cost, and in-app purchase details, replicating information visible on app stores in December 2024. Label cards present information exactly as shown on privacy nutrition labels or data safety labels, detailing data types and uses. Label cards do not include any information about the app to which they correspond; each label card is marked with a code to verify matches.

The nine apps include two social media apps (Facebook and TikTok), two music apps (Spotify and Apple/Samsung Music), two game apps (Chess.com and Candy Crush), two health and fitness apps (Strava and Nuffield Health Virtual GP), and one note-taking app (Notes/Samsung Notes). Two of the apps are pre-installed (Samsung Notes/Music and Notes/Apple Music); the others are all downloadable from the respective app stores. Apps were selected to cover multiple domains. We then chose popular apps within each domain. Figure 1 and Figure 2 show examples of the app cards and their corresponding label cards.

Due to the prominence of ***See details*** hyperlink on the DSL, the Android version of the tool had an additional set of nine "detail" cards, each linked to its primary label via the same code. These mirror the additional information accessed by clicking the hyperlink online. Participants were not informed about these cards and were only shown them if they asked for more details or questioned what the ***See details*** link was. With previous use of the tool (to facilitate conversation) there had never been a request for an extended version of the for the iOS labels, so we did not include this. However, definitions available through the ***Learn More*** link on the iOS labels were accessible through a set of definition booklets. These booklets provided definitions for all the terms used on the labels and could be shown or read to participants. The definitions were copied *ad verbatim* from definitions available online in December 2024 from click through links on the privacy labels. The availability of these definition booklets was not explicitly communicated, however, if participants asked the meaning of a term during the study, they were used to provide the definition. Though the task in this methodology differs from any real task that users would undertake as part of the app selection process, it presents an opportunity for users to engage with actual labels in the context of apps and requires them to think of app data-handling practices. As such, the tool has face validity.

3.3 Recruitment and Compensation

Participants were recruited using a snowball sampling method [4] by sharing a recruitment poster via social media and our networks, as well as through word of mouth. Snowball sampling was a deliberate choice, as the study aimed to gain in-depth insights into how app users interpret labels, rather than produce numerically representative findings. Eligibility criteria required participants to be over 18 years old and

smartphone users. Participants were compensated with £15 online shopping vouchers.

3.4 Pilot Study

Prior to launching the main study, we piloted it with three participants to test the study flow and check timings. None of the pilot sessions were recorded and no data was collected. Based on pilot observations, we developed a short script to guide participants through the study process to ensure that participants were primed consistently.

3.5 Study Procedure

The study was conducted through one-on-one, in-person sessions. While the participant information sheet (PIS) and consent form were emailed beforehand, printed copies were also available at each session. The four-part structure of the study - comprising an introduction, Activity One, Activity Two and a debriefing - was communicated in alignment with the PIS. Once written consent was obtained, audio recording began. The order of the two core activities was randomised (determined by a die roll) to control for potential order effects.

Following a standardised script to introduce the activity, the primary researcher asked participants if they had prior exposure to the labels and invited questions. The researcher then remained silent, intervening only to answer specific participant questions or to prompt a continued think-aloud (e.g., 'Don't forget to think aloud'). Upon completing the task, participants were informed of their initial score (e.g., 'You got three correct'), which was recorded by the researcher. Participants were then offered a second attempt to resolve any incorrect pairings, after which all correct pairings were revealed sequentially (e.g., 'Facebook is 4'). Depending on their level of engagement, participants either reflected spontaneously as each match was revealed or did so at the conclusion of the activity with all correct pairs displayed. If a participant did not offer unsolicited reflections, the researcher prompted them to think aloud. Following the second activity, participants completed a brief demographic questionnaire capturing age, gender, education, phone type, and app usage. Finally, they were thanked and asked if they wished to receive updates on the study's outcomes.

3.6 Data Analysis

3.6.1 Thematic Analysis

Thematic analysis was employed to analyse the qualitative data, following Braun and Clarke's evolved six-stage process. First, the primary coder familiarised themselves with the data, listening to, transcribing, and anonymising all audio recordings while taking initial notes (stage one). Initial codes were generated on the first reading of each transcript. Both existing and new codes were generated by re-examining the initial notes and re-listening to the audio. Data collection continued until no new codes were emerging ($n = 11$) deemed to be saturation point (stage two). Once all transcripts were initially coded, codes were sorted into groups to generate themes (stage three). At this stage, five additional sessions were conducted to validate saturation. All authors reviewed, discussed, and refined themes through multiple iterations (stage four). After five iterations, themes were defined, named, and organised in the codebook–Appendix A (stage five). The codes were organised using a thematic map (Appendix B). The findings from our final analysis are detailed below (stage six).

Given the unstructured, exploratory study design, our coding approach aligned with Braun and Clarke's reflexive thematic analysis [6, 7] and a constructivist approach [28]. We therefore did not calculate inter-coder agreement, as our goal was not to establish coding reliability through consensus metrics, but to develop interpretive themes through reflexive engagement with the data. The second researcher coded one script at the beginning to ensure aligned interpretation, but all the researchers ($n = 3$) discussed codes and evolving themes iteratively throughout to enrich the analysis.

3.6.2 Participants and Study Conditions

Our sample size ($n = 16$) was determined by the thematic saturation. Limited demographic information was collected as, owing to the qualitative methodology and sample size, no inferences can be made between demographic characteristics and the findings. However, some information was collected to contextualise the findings. Participant's ages ranged from the 18-24 to 55-64: 18-24 (5), 25-34 (2), 35-44 (4), 45-54 (4), 55-64 (1). Nine participants were female, seven male; the majority were educated to degree level (2 students, 14 graduate professionals). Twelve participants had devices running iOS, five Android (one participant had both). All participants used apps. Participants were not screened as the selection criteria was not platform dependent. Six participants started with the Android activity first. No participant reported having prior experience using privacy labels. Score for correct matches was on average 2.3/9 (37/144) for PNL and 3.1/9 (45/144) for the DSL. The average study duration was 39.3 minutes.

3.7 Limitations

As a qualitative study using thematic analysis, our findings are contextual and offer insights into a limited number of participants' understanding. They are therefore non-generalisable [46]. Limitations to our snowball sampling approach include homophily bias [35], restricted demographics [4], and potential researcher-acquaintance bias [17]. To mitigate single-source bias we used hybrid recruitment combining advertising and referrals.

We also considered app familiarity in the study design as a potential limitation. However, pilot testing and prior use of

activity suggested that app familiarity did not always translate to accurate matching, as participants still struggled with recognisable apps. To check for this we recorded app usage. Descriptively, familiarity did not consistently correspond to higher matching accuracy: while some familiar apps had relatively higher correct matches (Facebook 9/32 correct, Notes 15/32 correct), Spotify (reported to be used most frequently) was only matched correctly 2/32 possible times. None-theless app familiarity is a possible limitation of the study design.

4 Findings

The findings section is arranged in three parts, describing and providing evidence for each of the three themes generated by thematic analysis. A summary of the themes, sub-themes and codes is listed in Table 1. The full codebook and thematic map are available in Appendix A and B.

Table 1: Themes, sub-themes and codes per sub-theme.

Theme	Sub-themes	Codes
Understanding	Genre-Driven Contextual Expectations	6
	Operational and Feature-Linked Functional Expectations	5
	Perceived Rationale for Data Collection	4
	External Drivers of Mental Models	4
	User Comprehension of Technical Terminology	1
Salience	Visual Salience and Attentional Focus	4
	Observed Differences Between Label Types	4
Impact	Affective and Emotional Response	5
	Perceived Behavioural Impact	1
	Societal Implications	1

4.1 Understanding

This theme *Understanding* relates to RQ1: *What do users understand about privacy labels when they are asked to use them?* Our analysis distinguishes between users' mental models of app data-handling practices in general, and their direct semantic understanding of the privacy label interface itself.

With respect to users' mental models of app data-handling practices, we find that on a macro-level users possess a realistic, tacit understanding of the modern app economy, recognising that apps must necessarily harvest and process data to support core functions and drive monetisation strategies (*4.1.3 Perceived Rationales for Data Collection*). On a micro-structural level however, mental models are incomplete or inaccurate, with users demonstrating a lack of clarity regarding how fundamental data-handling practices map on to front-end features such as in-app purchases (*4.1.2 Operational and Feature-Linked Functional Expectations*). This systemic deficit is further illustrated by our latent analysis, which reveals evidence of the "halo effect" [8]. Rather than assessing data practices objectively, users' mental models are systematically biased by external heuristics like app genre or brand reputation (*4.1.4 External Drivers of Mental Models*) which in turn lead to users holding common app-genre-specific assumptions about data-handling practices (*4.1.1 Genre-Driven Contextual Expectations)*.

This underlying conceptual deficit directly impairs privacy label understanding (*4.1.5 Understanding of Technical Terms*). Instead of bridging these fragmented mental models and illuminating real-world data tracking, the privacy labels create a "transparency gap." Because the interface relies on abstract terminology and homogeneous structures, it fails to clarify how a user's data-sharing tolerance aligns with actual app behaviour. Critically, rather than improving user comprehension of app data-handling ecosystems, the labels actively compound existing confusion—fundamentally undermining their utility during active choice selection behaviour.

4.1.1 Genre-Driven Contextual Expectations

Our analysis highlights user assumptions regarding app data handling. These assumptions reflect pre-existing expectations revealed during the matching activity, rather than direct reactions to actual disclosed data practices.

Social media apps collect 'everything' (13/16). Most participants assumed that Facebook and TikTok would collect '*everything*'. P1 observes: *'Everything. Whatever you do, everything is taken. I find Facebook to be the most, the most invasive actually. Nothing is hidden. Like they collect, I think they collect everything.'*

Games don't need much data (12/16). Another common assumption is that games do not collect much data. For example, P14 says: *'Candy Crush, not a lot of information you can get out of that... Chess won't collect very much data.'*

Location should only be collected for functional purposes (10/16). Users strongly expected location data collection to align strictly with core app functionality. Consequently, tracking within non-functional contexts, such as mobile games, caused significant surprise. P10 noted *'Location, though I don't understand. Well, obviously the label 7 and label 2 might not go with Chess and Candy Crush, but if they track your location, it's like, why, because you don't need it.'*

Only health-related apps should collect health and fitness data (7/16). When looking at the labels, participants assumed that only health-related apps would collect health and fitness and / or sensitive data. For example, P16 observes: *'That collects health data so I'll put it with a health app'*. This logical assumption — that health data collection is restricted to health apps — underpins a broader user belief that data collection must serve a functional purpose. Interestingly, while three apps in the study actually handled health and fitness data, participants only categorised two (Nuffield Health and

Strava) as health apps. Users commonly assumed that the third app collecting this data must be Facebook. For example, P1 says *'I think that the health and fitness is... it's something that would go only in these three apps and I have it in three cards... the Strava, the Nuffield and the Facebook.'*

Simple apps, like note-taking apps, don't need much data (10/16). Participants assumed that because notes had a simple, straightforward function, it would not need to handle much data. For example, P14 says:*'I'm looking for the least information collected on them for the Apple Notes because I don't think they can have much information on there.'*

Good apps collect data more responsibly (8/16). Apps deemed to be 'good' included the Nuffield Health Virtual GP and the Chess.com apps. For both apps, there was an underlying assumption that because they were 'good', they would act more responsibly in their data collection practices. P10 observes: *'Because it's a medical app and I'm just assuming it's more trustworthy based on the fact that it's sort of a GP, so maybe they don't collect, maybe they collect only the data that they do actually need, which I guess then is more comforting. Rather than just like collecting everything that they can or just because they can.'*

4.1.2 Operational and Feature-Linked Functional Expectations

Confusion about tracked, linked, not-linked and shared data (11/16). Many (not all) participants held incomplete mental models of how data was used. For example, P3 understood the terms *tracking* and *sharing* to mean different things, demonstrating a fundamental lack of understanding of app data-handling practices generally. P3 reads *'This app share these data types with third parties.* Then comments *'I mean, no idea who those third parties are. It doesn't say anything about...maybe that's the tracking bits.... It doesn't say anything about tracking on these ones. I think it was because it used the word track on this one (PNL), yeah. So it says: data used to track you, whereas this one (DSL) the language, if that's even the same, I don't know if it is, but it says this app may share these data types with third parties. I mean, I don't really want people to share my data, but it sounds a bit less scary than track.'* The terms linked and tracked were also confused with P8 reading *Data linked to me* then commenting *'So it's just linking it to me, but they're not sharing it, but they're tracking it.'*

Moreover, given the ostensible definition of *identifiers* as something that identifies you, labels listing *identifiers* under *Data Not Linked to You* also caused confusion. P10 observes: *'Identifiers. It's actually saying it's identifying you and it's saying, putting it under data not linked to you, which is interesting.'* Similarly, P10 could not understand how *usage data* could be listed as *Data Not Linked to You*, as for them how they used their data was intrinsically linked to them.

In App Purchases, and the interplay between financial data, purchases and subscriptions (14/16). When participants considered which apps may collect "financial information" or "purchase history", it became apparent that these two terms are often confused or conflated. Participants assumed that purchase data was only relevant for apps in which you could purchase things. P16 says: *'Purchases. Presumably this is something which sells you something?'* The term "purchases" was misunderstood, with P12 commenting *'Well I feel like [the label for] Candy Crush must have purchases though, because you can purchase stuff on it'*. P3 illustrated the opacity: *'But I think generally it's just quite opaque really, and it doesn't explain why it needs this information and when it says things like purchase history, I'm questioning is that my whole purchase history? Can it see beyond what I'm purchasing in the app? It doesn't say in app purchase history.'*

Participants assumed that "financial data" could only be collected by apps with in-app purchases. P10 says: *'I'm thinking of an app where you might purchase... Now I'm thinking G might go with Chess because I feel like, the chess one, I think from my current understanding is you have In-App purchases to unlock levels or lessons. I'm seeing if any others mention financial information. And none others do, which I find quite surprising. I might put G with TikTok, so I think TikTok has a feature where there's like a TikTok shop now. And that's based on the fact that G mentions financial information.'*

Relationship Between Account and App (7/16). There is confusion about the relationship between apps and accounts. For example, P9 says: *'Just out of interest, so if you had an account on Chess.com, but you don't have the app, if you just log into the website each time, can they collect all that information still or is this specific to the app?'* The mention of accounts on the labels further compounds this confusion. P4 says: *'OK, let's see: account deletion available. So you can delete your account. I don't know what account I'd be deleting.'* Users appear unsure if downloading an app requires creating an account. P10 says: *'Because these two are the only two which describe it as an app rather than an account, I'm thinking it's these two are more likely to be Candy Crush and Chess. Because those are more games. And I would think out of the ones laid out, they are more likely to be described as apps than a rather than an account.'*

Assumptions of Cross-Platform Consistency (4/16). Users assumed that the same app, hosted on different platforms, would collect and use data in the same way. For example, participants expected Strava to employ consistent data-handling across iOS and Android. P15 observes: *'I made an assumption that they would be almost the same, that the*

data that all the companies collect, no matter what platform or device platform, it would be a similar approach. That's strange that it would be different.' While this reflects a logically consistent mental model from a consumer perspective, recent empirical audits reveal that over 60% of identical apps exhibit clear discrepancies in their data collection disclosures across the iOS and Google Play marketplaces [40].

App with the same principle function should collect the same types of data (8/16). A commonly held assumption is that apps with the same function should have the same or very similar data-handling practices. P11 says: *'And then we're left with Facebook, TikTok, Spotify and Apple Music. Right. I'm assuming these would be slightly like two distinct groups and that the music apps would have similar properties and the social media apps would have similar properties.'*

4.1.3 Perceived Rationale for Data Collection

Privacy labels provide explicit reasons for data collection. PNL specify that data may be used for tracking, linked to you, or not linked to you. DSL state that "This app may collect these data types" and/or "This app may share these data types with third parties." Though often misunderstood, these terms signal that apps collect data for different reasons.

For App Functionality (15/16). The main reason participants think apps collect data is for the app to function. Participants understand that apps need data to operate. These findings correlate with the common assumptions noted earlier, for example, "location should only be collected for functional purposes", "Only Health-focused Apps should collect Health and Fitness Data" but extend to a more general acceptance that certain apps require specific data types to function. For example, P16 says: *'This seems to collect video or audio you create, so I'll put that one with TikTok'* and P3 says: *'Maybe search history, like what you searched for before, would be more relevant to Apple Music. They might want to give you suggestions based on artists you search or something.'*

Because they can (10/16). Participants also thought apps collected data, particularly social media apps, simply because they can. This is illustrated by P6 who says: *'I mean, I would guess Facebook too just because they're obsessed with data collection.'*

For Money (8/16). Further, participants thought that data was collected for money. P4 observes: *'If they share it, they can make money from it.'* Extended understanding of why or how apps make money from collecting data was less clear. For example, P15 observes: *'I'm sure there are third parties who are very interested in my health data - if it's lucrative I expect that they probably do it.'*

To Know More About You: Advertising and Surveillance (8/16). Some participants acknowledged that data was collected to know about them for targeted advertising. For example, P4 says: *'1-2-3-4 apps use identifiers to track you...so they just track you from website to website this is...(laughs)... interesting. And then they just link your data to you, so I suspect that these track you...I feel like those companies that track you want to...they're tracking you for a reason, to try and see what you do, what you like, to try and understand and or how to bring you more usefulness or get you to stay with them and keep using their services, for example, like through adverts and stuff like that.'* Two participants extended this notion of collecting data to know more about you for surveillance purposes.

4.1.4 External Drivers of Mental Models

Our analysis shows that reputation, brand image, and media portrayal form the baseline of user understanding, rather than digital literacy or formal education. This understanding is systematically biased by the halo (or reverse-halo) effect [8], where a user's prior impressions and brand perception dictate how they assume an app handles data.

Brand Image (11/16). Participants made assumptions about app data-handling practices based on pre-existing attitudes relating to the app's brand. This links to the "good apps collect data more responsibly" assumption. For example, P4 says: *'There's two Apples and a Spotify. I think they're probably a worst case one is gonna be an Apple app, just so they can further their monopoly anyway'*, exemplifying the halo effect - the systematic bias in user perception meaning that apps with good functions have good data-handling practices.

General Reputation (8/16). The general reputation of app also skewed user expectation: an app's perceived utility, worthiness, or highbrow reputation also invoked a halo effect, where users anticipated better data stewardship for these apps. Participants consistently assumed that apps designed for health benefits or intellectual merit were inherently more trustworthy than 'frivolous" or social media apps. For example, P8 says: *'Financial info, yeah, they take my bank details. Again, there's still this kind of thing that, you know, if it's a health app, then it can be trusted more than a gaming app or a social media app.'* P3 demonstrates this in the context of health apps:*'the ones you feel are kind of a bit more worthy, like Strava, I feel like that's something that's potentially designed to do me good so I don't feel so weird about potentially giving it my data, so much of my data. Whereas Candy Crush, which is sort of a bit of a frivolous, unnecessary thing, it feels wrong'* Moreover, due to chess's "more prestigious reputation", users assumed Chess.com would not collect excessive data. As P15 exemplified:*'For the chess one, I've put F. So they share some data, but it's quite low. Maybe my own*

internal snobbery about just being highfalutin and that they wouldn't be engaged in that.'

Media (5/16). Users also reference the media as a source of their knowledge about app data-handling practices. For example, P4 says: *'I guess, in the news there's a lot of talks where TikTok does take a lot of your data and I found that quite interesting.'*

Experience (6/16). Past user experiences — including using apps or managing data professionally — directly inform structural assumptions. For instance, P8 noted relying on *"past impressions with the apps"*, while P6 leveraged professional ad-buying experience with Meta: *'...one of the... better ad campaigns that we ran was based on data that we bought off WhatsApp. People have been talking about their pensions and we served ads based on that and it worked brilliantly...'*.

4.1.5 User Comprehension of Technical Terminology

Engaging with the activity revealed that participants often did not understand the meaning of some terms used on labels. In these situations they would either attempt to infer meaning or explicitly ask what specific terms meant. The most commonly misunderstood or questioned terms were: *browsing history* (6), *search history* (5), *identifiers* (5), *diagnostics* (4), *sensitive data* (5), *other data* (3), *other IDs* (3), *personal information* (2), *user content* (2), *usage data* (2), *probabilistic identifiers* (1), *financial info* (2) *crashing*, and, *in-app messages* (2). Definitions were specifically requested to clarify differences between data linked, not linked and tracked (3).

On each occasion, the definition was provided from the definition booklets. This sometimes helped clarify what was being collected, but usually it did not. For example, P10 asks: *'Diagnostics? What does it mean by diagnostics where it says data not linked to you?'* After the researcher reads them the definition, P10 replies: *'OK, that just seems like a lot of words which just doesn't really give much actual information.'*

4.2 Salience: Things People Notice on Labels

The theme *Salience* encompasses two sub-themes: *Visual Salience and Attentional Focus* and *Observed Differences Between Label Types*. Even when given time and explicitly prompted to review privacy labels, participants rarely read them line-by-line. Instead, they relied on heuristic reasoning, combining pre-existing assumptions with interface features—such as list lengths or the volume of data types—to make sense of disclosures. This behaviour suggests current designs fail in practice because users do not seek a structural understanding of data-handling practices; rather, they prefer clear heuristic cues to guide their choices.

4.2.1 Visual Salience and Attentional Focus

This sub-theme collates codes relating to aspects of the privacy labels that were salient to them.

Gross Data Quantity (16/16). PNL and DSL formats differ structurally: PNL presents data types in soft-edged tiles arranged by use, accompanied by icons (see Figure 1), while DSL organises parameters within a single frame using descriptive text and numerical metrics (see Figure 2). Rather than parsing specific data types, participants systematically evaluated gross data quantity — relying on PNL list length or DSL numerical counts — to gauge data handling via a heuristic "more-or-less" approach. For example, P13 noted: *'...my strategy is looking at... the numbers [and] matching it to... the amount of collection of data I'd assume the app would like take...'*. Similarly, P15 bypassed granular details to look at numeric volume: *'...they look so similar... So there's seven types of data. There's twelve...'* P5 summarised this cognitive strategy as a "quick T-shirt sizing" model rooted in baseline genre assumptions: *'...which I would expect to have more and or more sensitive data versus less... Samsung note [less]... Music [middle]... Facebook [more]...'*.

Salient Data Types (13/16). Users leverage certain key terms to navigate the activity. P2 articulates this thought process: *'The way I would categorise this is maybe look at the information it's gathering and then kind of try to link it. For example, some of these I see there's location as one of the data points that is collected. So maybe like the fitness tracking app.'* This strategic approach allowed us to record which data types were most salient to the user. The most commonly leveraged data types were *health and fitness*, *financial information* and *location*. These terms may have been the most salient because they are common everyday terms and participants assumed complete understanding of them.

Visual Uniformity (8/16). A common observation was that the labels all look the same. Standardisation is useful for auditing [24], but the current form may habituate users so that they stop noticing aspects of the label. P15 says: *'But the wording of these two data safety cards looks nigh on identical.'*

Labels as Opaque, Vague, Confusing: Lacking Transparency (12/16). Participants often commented on the use of ambiguous terminology, noticing ambiguous terms like "may". This lack of transparency creates an inherent sense of mistrust, with participants anticipating that something may be intentionally hidden. P8 articulated this common concern, reading *'This app 'may' share these'* and then remarking: *'very vague. And confusing, almost intentionally confusing actually.'* This is a common assertion - the specific terms *vague*, *hidden* or *opaque* were used by 11/16 participants. In

addition, participants equated simplicity with transparency, with P8 remarking: *'I assumed that the most simple card that had less aspects to it less points, so its more ethical, more transparent, therefore more kind of safer.'*

4.2.2 Observed Differences Between Label Types

Although not analysed for statistical significance, there were more correct matches for the DSL (41) than the PNL (34). Although the majority of participants were Apple users, none had used privacy labels prior to this study, meaning that they were not biased by previous experience. All participants made direct comparisons between PNL and DSL (16/16), with terminological differences noted. Eleven participants (11/16) commented on DSL-specific features like account deletion information. Three participants (3/16) explicitly observed the use of symbols on PNL. Five participants (5/16) commented on the nutrition analogy.

4.3 Impact

Analysing participants' interactions during the matching activity and their reflections on the correct pairings allowed us to observe their reactions to disclosed data-handling practices in context. This provided clear indicators of how increased disclosure awareness might shape user perceptions. These findings are organised into three sub-themes: *Affective and Emotional Response*, *Perceived Behavioural Impact*, and *Societal Implications*.

4.3.1 Affective and Emotional Response

The language used during the activities was indicative of strong feelings about data-handling practices. This is a latent sub-theme, inferring meaning from context and the language used.

Negative Sentiment (14/16) Often the language used to describe data-handling practices had negative sentiment, leading us to infer that increased awareness of actual data-handling practices results in an increase in negative affect. These included words such as: invasive or intrusive (7), participants reported feeling sickened or described app data-handling practice as creepy or scary (7), and uncomfortable with the collection of personal and sensitive data (6).

Sense of Ownership (9/16). The language used to describe data-handling practices implies a sense of data ownership. For example, data was regularly described as being 'given away', 'taken', or even 'stolen'. Users appear to recognise an inherent value in their data, albeit framed negatively. P3 says: *'I didn't realise I was giving away so much of my data for free by downloading.'* While there was a tacit understanding that apps sometimes collect data to make money, for Spotify - where a subscription can be charged and the app appears to collect significant amounts of data - the enhanced understanding of the app data-handling practice caused annoyance. P6 says: *'I'm more annoyed with it about the Spotify because you pay for it. And I think well you're paying for it, they should be behaving themselves and keeping their, you know. I think you either get the money or the data... taking both seems a bit greedy.'*

Shock, Surprise, Disbelief (13/16). When asked to reflect on correct pairings, the most common reaction was surprise, shock or disbelief. For example, P11 says: *'Oh wow. Yeah, that's quite a surprise. Didn't realise that Chess tracked that. And then TikTok.'* The word 'surprise' was used 38 times by 9 participants. Words like 'insane' (P4), 'outrageous' (P6), 'crazy' (P4, P12, P13) were commonly used to describe how participants felt when actual app data-handling practice were revealed. Despite expecting extensive data collection, participants were surprised by actual practices.

Trust (10/16) Whether through a perceived lack of transparency (observed as a salient feature of labels) or in relation to brand or app, trust or the trustworthiness of apps was a constant theme. Trust was used in a negative sense, with participants articulating feelings of mistrust. P8 says *'I don't like the whole setup, I don't really, yeah. And that I would expect it to be reaping... Reaping I think is the work, grabbing lots of data for me. I mistrust it anyway.'* Similarly, P5 comments *"Tiktok - I'm deeply distrustful of it."* This highlights the importance of trust to users.

User Apathy, Overwhelm (10/16) A sense of apathy emerges from cognitive overload and confusion, rendering privacy labels difficult to use at a glance. Participants frequently found the volume of technical, unfamiliar information overwhelming. P3 noted it is *"hard to sort of process this information at a glance,"* while P8 stated, *"My mind wants to switch off. It's basically kind of incomprehensible and overwhelming."* As a result, when confronted with technical terms, users often default to installing the app regardless of privacy risks. As P10 explained *"I'm not too sure what an IMEI number or device ID is... I would probably read that and just be like, OK, I don't really understand, but I kind of want to use Spotify anyway and still get it...."*

4.3.2 Perceived Behavioural Impact

(8/16). Privacy labels aim to inform users to facilitate decision-making. This sub-theme explores how engaging with labels alongside apps provides users with advanced knowledge of data practices that could shape their future installation and usage behaviours. Half of the participants organically reflected on their future app selection or download behaviour. P13 says: *'The thing is when I download stuff, I don't even care about*

looking at like what information it collects, whatever. So obviously like, that's why I find it like so hard because I didn't think like, I don't know. I don't really care if, like they take my location or like they take my contacts. You know, I just want to use the app.'

Increased awareness of app data-handling practice also caused participants to think about deleting apps. P9 says: *'That's really bad. That's what you put in your notes. The user content, that's what you're writing in your notes, which is your kind of personal to do lists or like, and potentially kind of very private information that's that. I think I might stop using Notes.'* This provides some evidence that increased understanding of app data-handling practice can impact user app selection behaviour.

4.3.3 Societal Implications

(6/16). This sub-theme highlights participant concerns regarding the broader societal impacts of data collection, emphasising the need for collective action over individual effort. For instance, P10 notes that individual resistance feels futile without widespread behavioural change P10 : *'I feel like unless it's highlighted more and a lot of people stop using it ... one person that's not really going to make it, they're not going to change their data safety label and not share the personal information with third parties just because one person's like, I'm not going to download it unless you don't share my personal information with third parties.'*

Participants also questioned the motives behind societal-scale data harvesting. P16 expressed willingness to share data for public good (e.g., government health interventions) but worried about exploitation by "corporate money-making entities." *'I would like to think, that as much energy was taken to harvest our data to affect positive change, if the government looking to target health interventions then I wouldn't mind, but these are all corporate money making entities that are only going to use it to target people in a different way.'* Furthermore, this societal concern extended to vulnerable populations, with P8 identifying heightened risks in children's applications: *"That's the child's game thing... they use a lot of data tracking me... I mean location is kind of weird. Why are they? They need to know where I am. That's really dangerous."*

5 Discussion

5.1 Beyond Recall: Capturing Interpretation and Affective Impact

Our applied-comprehension approach reveals how users actually synthesise information when navigating privacy label interfaces. This is a dimension that traditional usability metrics, such as information-finding speed, text readability, and recall, are not designed to capture [22, 48, 49]. This shift in method also expands the dimensionality of how label exposure affects users. Existing baselines largely measure shifts in perceived security risk and willingness to install [2, 15]; our findings show that encountering accurate data-handling disclosures triggers consequences that extend well beyond functional consumer choice. In particular, participants exhibited clear affective and emotional responses such as negative affect, reduced trust, heightened apathy, shock, and disbelief. This suggests that label exposure is not merely an informational event but an emotionally consequential one. This is vital because it implies that evaluating labels purely on comprehension or recall risks overlooking a significant component of how users actually experience and respond to disclosure. Our methodology facilitates a pivot from recall to active application, allowing us to reframe the current understanding of how design choices shape consumer assumptions. This reframing also surfaces a deeper limitation in existing evaluation frameworks. For instance, approaches such as Zhang et al.'s [48] import criteria designed for compliance-driven privacy consent notices, where the goal is simply acknowledgement of disclosure. Our data supports the critique that such metrics do not adequately translate to choice-driven label ecosystems, where the goal is not simply acknowledgement but informed, comparative decision-making.

5.2 Current Privacy Labels Fail to Meet their Informational Promise

Our findings both validate and expand the existing literature on privacy label usability. While privacy labels inherently improve upon dense, legalistic privacy policies [22, 23], they still consistently result in widespread user misunderstanding and suffer from pervasive design and structural limitations [25, 33, 48]. Our work aligns with this baseline consensus, confirming that privacy labels have yet to fully realize their potential to inform user choice or support preference-based market navigation [21]. However, whereas prior literature largely attributes these shortcomings to isolated factors such as the poor comprehension of specific technical terms [25, 33] or sub-optimal layout formats [49], our empirical takeaways highlight a more fundamental gap. Moreover, we show that even when the structural barriers highlighted by Cranor [12] are neutralised (by removing the need to find the labels and giving users unlimited time to use them), labels still fail to facilitate comprehensive reading. Overall, we demonstrate that semantic misunderstandings of individual terms are in fact symptomatic of deeper, structural misunderstanding: namely, a systemic deficit in users' conceptual mapping of the overarching data ecosystem, and a historical absence of label designs grounded in actual user behaviour. This comprehension gap is further compounded by research highlighting that a significant percentage of apps are in fact fundamentally inaccurate or inconsistent across platforms [26, 40]. This means that even a fully informed user would be navigating a

landscape of unreliable information.

5.3 Heuristic Engagement as Rational Adaptation to Privacy Complexity

This shift toward capturing how users actually engage with labels explicitly surfaces the heuristic shortcuts users deploy when interacting with privacy labels, providing robust empirical evidence to support heuristic-led design paradigms. Research on heuristic decision-making demonstrates that "fast and frugal" decisions primarily depend on a limited set of highly salient factors [18]. Under real-world constraints, decision-making is naturally fast and frugal: because many users are unwilling or unable to expend the cognitive energy required to process exhaustive disclosures, they default to heuristic cues as a rational coping mechanism for managing the cognitive load associated with privacy fatigue [11]. We observe two primary heuristic strategies at play. First, users rely on a *halo effect* to judge data safety, assuming that apps with positive brand reputations or intrinsically worthy functions must inherently employ responsible data practices. Second, users project pre-existing mental models of data handling whether accurate, incomplete, or flawed onto label content, consistently prioritising gross data quantity as a proxy for risk over nuanced data classifications. Critically, these strategies do not reflect a failure of user attention or literacy but they represent a rational adaptive response to poorly calibrated design. Users are not ignoring privacy labels; they are reading them differently than designers assume.

5.4 Toward Heuristic-Led Privacy Label Design

This re-framing carries direct and concrete implications for label design. Current label formats implicitly assume an analytical reader who will systematically process each data category and arrive at informed conclusions. This is an assumption that our findings fundamentally do not support. A heuristic-led design paradigm would emphasise the signals users already rely on, making salient cues more legible and reliable rather than demanding engagement with granular detail. Future research should test whether portraying app data-handling practices numerically better aligns with observed heuristic strategies and more effectively supports consumer choice than existing categorical disclosure formats. For instance, future designs could surface aggregate data counts or other summary indicators that allow users to quickly assess relative privacy risk across apps. This approach maps to research which explores the potential of embedding privacy information in visual metaphors and analogies to support informed decision-making for cookie notices and other privacy disclosures [51]. Importantly, designing for heuristic engagement would also address structural barriers identified elsewhere in this research, particularly the visual homogeneity (i.e., "labels all looking the same") that makes labels difficult to distinguish from one another and the opacity introduced by ambiguous terminology. This approach treats these barriers not as isolated usability issues, but as part of a broader strategy for aligning label design with actual user behaviour.

Our research calls for a shift from exhaustive privacy disclosures to curated, accurate cognitive shortcuts. To this end, we offer the following recommendations:

1. Designers should identify and prioritise a small set of salient, theory-grounded heuristic cues to drive relevance. Labels should translate these few choice-critical metrics into immediate, legible interface elements that reflect the limits of real-world decision-making. By aligning visual design with empirical cognitive limitations, designers can ensure privacy labels operate as highly effective, behavioural decision aids.
2. Future labels should represent data flow itself rather than presenting isolated text classifications. For instance, visualising the pipeline from collection to transmission to downstream sharing. This could help users form a clearer mental model of otherwise abstract practices.
3. Researchers and practitioners should adopt a standard, multi-dimensional framework to evaluate label effectiveness, moving beyond traditional variables like reading time or word recall to assess decision-making support and the reduction of negative outcomes such as privacy fatigue and negative affect.

6 Conclusion

There is an ethical imperative to ensure that labels effectively inform user choice so individuals can navigate the app ecosystem safely and in line with their preferences. When labels fail to provide clear and accurate guidance, user autonomy is compromised and the purpose of labels is defeated. Misunderstood or inaccurate disclosures may also create additional security and privacy risks, encouraging behaviours such as excessive data sharing or neglect of basic safeguards. For privacy labels to be more than a PR exercise, they must do more than disclosing information, they must efficiently highlight the most salient aspects of app data practices and support informed decision making. These findings provide new insights for developing more effective user-centric privacy labels.

Acknowledgments

We would like to thank all the participants would took part in our study and the reviewers who provided useful and constructive feedback. Sophia Walsh is a PhD student supported by EPSRC CDT TIPS-at-Scale EP/S022465/1. Ramokapane's time was supported by REPHRAIN EP/V011189/1 and Equitable Privacy EP/W025361/1.

References

[1] Apple Inc. App privacy details - App Store. https://developer.apple.com/app-store/app-privacy-details/, 2020. Accessed: 2026-02-17.

[2] David G Balash, Mir Masood Ali, Chris Kanich, and Adam J Aviv. " i would not install an app with this label": Privacy label impact on risk perception and willingness to install {iOS} apps. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*, pages 413–432, 2024.

[3] Bishnu Bhusal, Yuanye Ma, and Rohit Chadha. Privacy nutrition labels: Promise, practice, and paradoxes in communicating privacy. In *International Conference on Human-Computer Interaction*, pages 18–28. Springer, 2025.

[4] Patrick Biernacki and Dan Waldorf. Snowball sampling: Problems and techniques of chain referral sampling. *Sociological methods & research*, 10(2):141–163, 1981.

[5] T. Boren and J. Ramey. Thinking aloud: reconciling theory and practice. *IEEE Transactions on Professional Communication*, 43(3):261–278, 2000.

[6] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[7] Virginia Braun and Victoria Clarke. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health*, 11(4):589–597, 2019.

[8] Paul F Burke, Grahame Dowling, and Edward Wei. The relative impact of corporate reputation on consumer choice: beyond a halo effect. *Journal of Marketing Management*, 34(13-14):1227–1257, 2018.

[9] Peter J Caven, Shakthidhar Gopavaram, Jayati Dev, and L Jean Camp. Sok: anatomy of effective cybersecurity label development. *Available at SSRN 4591786*, 2023.

[10] Jiahong Chen. *Regulating Online Behavioural Advertising Through Data Protection Law*. Edward Elgar Publishing, 2021.

[11] Hanbyul Choi, Jonghwa Park, and Yoonhyuk Jung. The role of privacy fatigue in online privacy behavior. *Computers in Human Behavior*, 81:42–51, 2018.

[12] Lorrie Faith Cranor. Mobile-app privacy nutrition labels missing key ingredients for success. *Communications of the ACM*, 65(11):26–28, 2022.

[13] David Curry. App downloads data (2026), 2 2026. Updated: February 3, 2026.

[14] Pardis Emami-Naeini, Yuvraj Agarwal, Lorrie Faith Cranor, and Hanan Hibshi. Ask the experts: What should be on an iot privacy and security label? In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 447–464. IEEE, 2020.

[15] Pardis Emami-Naeini, Janarth Dheenadhayalan, Yuvraj Agarwal, and Lorrie Faith Cranor. Which privacy and security attributes most impact consumers' risk perception and willingness to purchase iot devices? In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 519–536. IEEE, 2021.

[16] Grace Fox, Theo Lynn, and Pierangelo Rosati. Enhancing consumer perceptions of privacy and trust: a GDPR label perspective. *Information Technology & People*, 35(8):181–204, December 2022.

[17] Alistair Geddes, Charlie Parker, and Sam Scott. When the snowball fails to roll and the use of 'horizontal'networking in qualitative social research. *International Journal of Social Research Methodology*, 21(3):347–358, 2018.

[18] Gerd Gigerenzer, Jochen Reb, and Shenghua Luan. Smart heuristics for individuals, teams, and organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, 9(1):171–198, 2022.

[19] Google Help. Google Help YouTube Channel, 2026. Official Google Support Channel.

[20] Johanna Johansen, Tore Pedersen, Simone Fischer-Hübner, Christian Johansen, Gerardo Schneider, Arnold Roosendaal, Harald Zwingelberg, Anders Jakob Sivesind, and Josef Noll. A multidisciplinary definition of privacy labels. *Information & Computer Security*, 30(3):452–469, 2022.

[21] Shane D Johnson, John M Blythe, Matthew Manning, and Gabriel TW Wong. The impact of iot security labelling on consumer product choice and willingness to pay. *PloS one*, 15(1):e0227800, 2020.

[22] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W Reeder. A nutrition label for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, pages 1–12, 2009.

[23] Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, and Lorrie Faith Cranor. Standardizing privacy notices: an online study of the nutrition label approach. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, pages 1573–1582, 2010.

[24] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI Conference on*

Human Factors in Computing Systems, CHI '13, page 3393–3402, New York, NY, USA, 2013. Association for Computing Machinery.

[25] Ishika Keswani, Kerick Walker, Adrian Clement, Eusila Kitur, Nannapas Wonghirundacha, Ryan Aubrey, Vivien Song, and Eleanor Birrell. User understandings of technical terms in app privacy labels. In *Twenty-First Symposium on Usable Privacy and Security (SOUPS 2025)*, pages 279–298, 2025.

[26] Rishabh Khandelwal, Asmit Nayak, Paul Chung, and Kassem Fawaz. Unpacking privacy labels: A measurement and developer perspective on google's data safety section. *arXiv preprint arXiv:2306.08111*, 2023.

[27] Rishabh Khandelwal, Asmit Nayak, Paul Chung, and Kassem Fawaz. Unpacking privacy labels: A measurement and developer perspective on google's data safety section. In *33rd USENIX Security Symposium (USENIX Security 24)*, pages 2831–2848, Philadelphia, PA, August 2024. USENIX Association.

[28] Charles Kivunja and Ahmed Bawa Kuyini. Understanding and applying research paradigms in educational contexts. *International Journal of higher education*, 6(5):26–41, 2017.

[29] Simon Koch, Malte Wessels, Benjamin Altpeter, Madita Olvermann, and Martin Johns. Keeping privacy labels honest. *Proceedings on Privacy Enhancing Technologies*, 2022.

[30] Konrad Kollnig, Anastasia Shuba, Max Van Kleek, Reuben Binns, and Nigel Shadbolt. Goodbye tracking? impact of ios app tracking transparency and privacy labels. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 508–520, 2022.

[31] Tianshi Li, Lorrie Faith Cranor, Yuvraj Agarwal, and Jason I Hong. Matcha: An ide plugin for creating accurate privacy nutrition labels. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(1):1–38, 2024.

[32] Tianshi Li, Kayla Reiman, Yuvraj Agarwal, Lorrie Faith Cranor, and Jason I. Hong. Understanding challenges for developers to create accurate privacy nutrition labels. In Simone D. J. Barbosa, Cliff Lampe, Caroline Appert, David A. Shamma, Steven Mark Drucker, Julie R. Williamson, and Koji Yatani, editors, *CHI '22: CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 29 April 2022 - 5 May 2022*, pages 588:1–588:24. ACM, 2022.

[33] Yanzi Lin, Jaideep Juneja, Eleanor Birrell, and Lorrie Faith Cranor. Data safety vs. app privacy: Comparing the usability of android and ios privacy labels. *Proceedings on Privacy Enhancing Technologies*, 2024.

[34] James McParlan and Dirk van der Linden. Privacy labels should go to the dogs. In *Proceedings of the Eight International Conference on Animal-Computer Interaction*, pages 1–10, 2021.

[35] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[36] David Nguyen and Marta Paczos. Measuring the economic value of data and cross-border data flows: A business perspective. Technical report, OECD Publishing, 2020.

[37] Milena Nikolić and Tijana Dabić. The bloom's taxonomy revisited in the context of online tools. In *Sinteza 2016-international scientific conference on ICT and E-business related research*, pages 315–320. Singidunum University, 2016.

[38] Martin T Orne. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. In *Sociological methods*, pages 279–299. Routledge, 2017.

[39] Kopo M Ramokapane, Maria Sameen, and Zaina Dkaidek. Inclusive internet of things privacy labels. *IEEE Security & Privacy*, 22(5):32–39, 2024.

[40] David Rodriguez, Akshath Jain, Jose M Del Alamo, and Norman Sadeh. Comparing privacy label disclosures of apps published in both the app store and google play stores. In *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 150–157. IEEE, 2023.

[41] Yun Shen and Pierre-Antoine Vervier. Iot security and privacy labels. In *Privacy Technologies and Policy: 7th Annual Privacy Forum, APF 2019, Rome, Italy, June 13–14, 2019, Proceedings 7*, pages 136–147. Springer, 2019.

[42] Meixue Si, Shidong Pan, Dianshu Liao, Xiaoyu Sun, Zhen Tao, Wenchang Shi, and Zhenchang Xing. A solution toward transparent and practical ai regulation: Privacy nutrition labels for open-source generative ai-based applications. *arXiv preprint arXiv:2407.15407*, 2024.

[43] Samuel Wairimu, Leonardo Horn Iwaya, Lothar Fritsch, and Stefan Lindskog. On the evaluation of privacy impact assessment and privacy risk assessment methodologies: A systematic literature review. *IEEE Access*, 12:19625–19650, 2024.

[44] Sophia Walsh, Lukasz Piwek, and Kopo Marvin Ramokapane. Three systemic problems with the privacy notice ecosystem. *Policy & Internet*, 18(2):e70042, 2026.

[45] Sophia Walsh and Kopo M. Ramokapane. App snap: A Versatile Tool for Engaging Children and Young People in Dialogue About Smartphone App Privacy. *Creating Engaging Security and Privacy Educational Interfaces for Educators and Families*, 2024.

[46] Robin Whittemore, Susan K Chase, and Carol Lynn Mandle. Validity in qualitative research. *Qualitative health research*, 11(4):522–537, 2001.

[47] Yue Xiao, Zhengyi Li, Yue Qin, Xiaolong Bai, Jiale Guan, Xiaojing Liao, and Luyi Xing. Lalaine: Measuring and characterizing Non-Compliance of apple privacy labels. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 1091–1108, Anaheim, CA, August 2023. USENIX Association.

[48] Shikun Zhang, Yuanyuan Feng, Yaxing Yao, Lorrie Faith Cranor, and Norman Sadeh. How usable are ios app privacy labels? *Proceedings on Privacy Enhancing Technologies*, 2022.

[49] Shikun Zhang, Lily Klucinec, Kyerra Norton, Norman Sadeh, and Lorrie Faith Cranor. Exploring ExpandableGrid designs to make iOS app privacy labels more usable. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*, pages 139–157, Philadelphia, PA, August 2024. USENIX Association.

[50] Shikun Zhang and Norman Sadeh. Do privacy labels answer users' privacy questions? In *Workshop on Usable Security and Privacy*, 2023.

[51] Verena Zimmermann, Adrienn Toth, Hannah Sievers, Linda Fanconi, Yanis Isenring, Mona Henz, Alina Stöver, and Nina Gerber. Let's get visual-testing visual analogies and metaphors for conveying privacy policies and data handling information. In *2025 IEEE Symposium on Security and Privacy (SP)*, pages 2303–2321. IEEE, 2025.

A Appendix: Codebook

Sub-Theme	Code	Code Description	Example(s)
THEME ONE: UNDERSTANDING			
Genre-Driven Contextual Expectations	Social media apps collect "everything"	Reference to Facebook or Tik Tok collecting lots of data, the most, 'everything'	"D will go to Tik Tok. Just because they collect in everything. Just after Facebook, of course."
	Games don't need much data	Reference to either game app (candy crush, chess) or games in general not needing location or questioning why games would need much data	"Wow. Quite small, superficial stuff, because again, it's a game but what you know, my mind is blown."
	Location should only be collected for functional purposes	When questioning why location would be collected when the collection of location data not clearly necessary for app functionality	"So maybe the easiest one to start with is Strava, because I feel like that uses the most physical location data, more than anything else."
	Only Health Focussed Apps should collect Health and Fitness data	Expressing assumptions about when health and fitness data should be collected. Also sensitive data should only be collected by health apps.	"That collects health data so I'll put it with a health app."
	Simple apps, like note taking apps, don't need much data	Reference to Notes not needing much data because simple function, reference to other simple apps/apps with simple functionality needing less data	"I guess the more simple the app, the more simple you actually expect the app to be, you assume the app to be, the less data you think it's robbing off, taking off you."
	Good apps collect data more responsibly	Reference to apps perceived to have a good or wholesome (GP, chess) being better or somehow more trustworthy	"And then and then I did for the chess one, I've put F. So they share some data, but it's quite low. Maybe my own internal snobbery about just being highfalutin and that they wouldn't be engaged in that"
Operational and Feature-Linked Functional Expectations	Confusion over how data is used: tracked, linked, not-linked and shared data	Expressing confusion / holding inaccurate or incomplete understanding with regard to data uses or between descriptions of data uses	"OK. What's the can you tell me? Like, what the difference is between data linked to you and then data not linked to you. Is it like these? Both sets are collected?"
	In App Purchases, and the interplay between financial data, purchases and subscriptions	Expressing confusion / holding inaccurate or incomplete understanding with regard financial or purchases data in relation to in-app or subscription	"And also it doesn't mention about financial information and it's a free app. Oh, actually, that that one doesn't mention financial oh, it says purchases. Oh, it does say purchases."
	Relationship Between Account and App	Expressing confusion / holding inaccurate or incomplete understanding with regard apps and accounts	"Contacts. I don't know why they need my contacts seeing as you have your own Spotify account names."
	Assumptions of Cross-Platform Consistency	Explicitly comparing same app on different OS and expressing assumption of same data practices	"But surely I have a higher chance of succeeding because you told me the answers for the first round, and I'm assuming that it's not going to be wildly different just because we're on a different platform."
	App with the same primary function should collect the same type of data	Comparing same app hosted on different OS	"Yeah, and I guess given that Apple music and Spotify, I would think would do a relatively similar thing, it looks like Spotify is more invasive in the data that it's monitoring slash harvesting, so I guess I'm a bit surprised by that."

Continued on next page...

Table 2 — Continued from previous page

Sub-Theme	Code	Code Description	Example(s)
Perceived Rationale for Data Collection	For App Functionality	Recognition that any data type is collected in order for app to function	"So you have health and fitness here that are linked to, your location, contacts. So this is giving me idea that's going to be about an app that has health information. So I have to in mind now I have the Nuffield Health or the Strava."
	Because they can	Implicit assertion that data collected for the sake of it, because it can, because that what apps do - without other reason	"I don't know what's in it for Samsung. If they're not collecting information where potentially they can."
	For Money	Reference to data collection being to make money or as part of business model	"So basically H and I are just saying - don't worry about us. We're not gonna take your details or any of your stuff at all. I think, I guess those are two apps which I'd assume are things which are not there to make money."
	To Know More About You, For Advertising and Surveillance	Reference to data being collected to learn about behaviour-both for OBA purposes but any other reason too	"TikTok is probably gonna use your health and fitness, but it could, because if they if they see your health and fitness and they see you're into your cardio or you're actually to health and fitness. they can just shove as directly to you in terms of that because I get a lot of TikTok ads."
External Drivers of Mental Models	Brand Image	Recognising that data collection practices are somehow related to brand - but now other reason	"I would like to connect this card that has the purchases and identifiers as data used to track you and the health and fitness data linked to you, I would like to put it with the Apple music for some reason."
	General Reputation Media Experience	Relates to more general reputation. Similar to 'good apps collect data more responsibility'. When understanding informed by basic assumption of reputation not directly linked to brand. Reference of media as source of knowledge Refer to experience of app use as reason assumption exists - also experience gained in professional capacity	"And then and then I did for the chess one, I've put F. So they share some data, but it's quite low. Maybe my own internal snobbery about just being highfalutin and that they wouldn't be engaged in that. But that's a ridiculous thing to base it on. For TikTok I put card C, because they do share a lot of your data - and I'm going on what's available in popular culture or in social media Contacts for like when - because sometimes, when you're on your full view page it might be like oh people you might know whatever.... So obviously they're using like your contacts to assume that you might know the person."
User Comprehension of Technical Terminology		Either asking for clarification, for a definition or lack of understanding implicit	"What's the different between browsing history and search history?"
THEME TWO: SALIENCE			
Visual Salience and Attentional Focus	Gross Data Quantity	Users referring to blocks of data as lists, number, quantifying amount rather than looking at detail.	"A lot of these have the same data collected; location personal info. so it makes it slightly harder to narrow down. E and B look fairly similar except B collects location personal and seven other data types, whereas E only collects five other types."
	Salient Data Types	Terms that jump to the user / appear salient/are explicitly noticed as they use the label	"I can't believe like how many ask for your location"
	Visual Uniformity	Observations that the labels all look the same	"Wow. They all look the same."

Continued on next page...

Table 2 — Continued from previous page

Sub-Theme	Code	Code Description	Example(s)
	Labels as Opaque, Vague, Confusing: Lacking Transparency	Refer to labels as vague, opaque, hiding, secretive, lacking transparency	"I'll consider myself to be a fairly kind of digitally savvy person. And I find this quite opaque"
Observed Differences Between Label Types	Making direct comparisons	Directly comparing the labels	"I guess like the difference of like the layout of like information. I think Apple. I think the words is easy to like follow with Android. But I think like Apple is, I know more specific like it lays out more"
	Symbols on PNL	Mention of icons or symbols	"Oh, you get little pictures."
	Difference with DSL including additional layer and account deletion	Explicitly noticing how DSL different; commenting on additional aspects of DSL; deletion, independent security etc.	"I guess the first thing that jumps out is more my explanation of where data is collected stored, how it's encrypted and if you can delete it or not."
	The Nutrition Analogy	Commenting on the use of the work nutrition	"If we're using nutritional, I suppose they are healthier."
THEME THREE: IMPACT			
Affective and Emotional Response	Negative Sentiment	Use of negative language relating to feelings: words include intrusive, invasive, disappointing, creepy, fear, sickening, or express worry	"I think the only reflection I have is like how scary it is."
	Sense of ownership	Uses terminology to do with ownership including stealing, taking, giving away data. Also reference to greed.	"I didn't realise I was giving away so much of my data for free, potentially by date, by downloading."
	Shock, surprise and disbelief at data handling practices Trust	Uses terms like shock or surprise or disbelief in response to app handling practices. Trust- or lack of explicit or implicit	"Yeah, I mean, they're all a little bit outrageous, to be honest. Again, there's still this kind of thing that, you know, if it's a health app, then it can be trusted more than a gaming app or a social media app."
	Hard Work To Use / Overwhelming	Infer that using labels hard work-cognitively demanding either overwhelm or confusion	"Oh wait app activity, app info performance and device. Fair enough. Well, I'm not like when you're telling me that I'm not like, oh, my God. I'm kind of just like, I don't know, my head hurts."
Perceived Behavioural Impact		Mention future app selection behaviour	"it's just now makes me go back home and check all of these on my phone."
Societal Implications		Reference to social impact from collective data collection	"I'm not aware of what the real world consequences might be that might directly impact on me. Is it more what they do with the information they're taking from society?"

B Appendix: Thematic Map

