Blissful Ignorance: A Visual Study of Perceived Privacy and Personal Data Flows

May 15, 2026

Adrian Vincentz Otto - ADOT@itu.dk - 22553
Frederik Alexander Schaufuss Borg - FBOR@itu.dk - 22608

BA project - BIBAPRO1PE - Spring 2026

1 Abstract

This thesis explores how non-technical users perceive the flow of their personal data in everyday digital interactions, and whether a structured visual language can help make the gap between that perception and the regulatory and technical reality more concrete and comparable. We develop a four-axis data flow card system (What, From, How/Why, Time) connected by bidirectional arrows on a layered topology, with the four axes loosely mirroring what controllers must already document under GDPR. The instrument is tested across seven semi-structured intercept interviews with non-technical participants, organised around three everyday cases anchored in Danish Data Protection Agency rulings: cookies, login, and chat. Within this small and demographically narrow sample, the most consistent pattern in the user pillar appeared to be not misinformation, but an articulated reluctance to engage with the question at all. Most participants reasoned about the first recipient of their data and went no further: only four of seven extended the model to a second-layer recipient, and no participant placed Meta in the Cookies case, even where they had marked *Received from third party* on the primary card. Where the technical chain was opaque, brand trust seemed to stand in for knowledge. One participant described the stance in their own words as *lykkelig uvidenhed*, blissful ignorance. The over-selection visible on the primary card may indicate confirmation-driven uncertainty rather than informed concern, while the absence of second-layer cards is at least suggestive of where participants' mental models tend to stop. As a proof of concept, the visual language appears workable as a research instrument for surfacing this kind of gap, though the findings should be read as exploratory rather than conclusive.

Contents

1 Abstract **1**

2 Introduction **3**

3 Literature Review **4**
3.1 Privacy as a Contested Concept 4
3.2 The Privacy Paradox and User Behaviour 5
3.3 Privacy Visualisation and Communication 6
3.4 The Gap This Project Addresses 7

4 Methodology **8**
4.1 Development of the Visual Language 8
4.1.1 What was intentionally cut 9
4.2 Case Construction from DDPA Rulings 10
4.2.1 Case A: Cookies 10
4.2.2 Case B: Login 10
4.2.3 Case C: Chat 10
4.3 Qualitative Interview Design 11
4.4 Analytical Framework 11

5 Results and Analysis **12**
5.1 Case Analysis 13
5.1.1 Case A: Cookies 13
5.1.2 Case B: Login 14
5.1.3 Case C: Chat 15
5.2 Cross-Case Patterns 16
5.2.1 Blissful Ignorance: Not-Knowing as the Default Mode 16
5.3 Data Credibility 19

6 Discussion **20**
6.1 Over-selection as a Sign of Not Knowing 20
6.2 What the Visual Language Reveals and What It Does Not 20

7 Limitations **21**

8 Implications and Future Work **22**

9 Conclusion **24**

References **25**

Appendices **27**

2 Introduction

Personal computers mediate most of our interactions with digital services. Browsing, using applications, and relying on cloud-based features all entail continuous processing of personal data, much of which happens in the background and is not visible to the people whose data it is. Privacy in this setting is shaped by three forces at once: the technical reality of how data is handled, the regulatory definition encoded in instruments such as the GDPR, and the everyday understanding that non-technical users build from the interfaces in front of them. Settings panels, permission prompts, and consent dialogs are the principal channel through which privacy is communicated, and they rarely reflect the underlying processing in full, particularly as automated and AI-supported systems take on more of that processing. In this thesis, we therefore ask:

How do non-technical users perceive the flow of their personal data in everyday digital interactions, and can a structured visual language make the gap between that perception and regulatory and technical reality concrete and comparable?

The thesis serves as a proof of concept that examines how far non-technical users' assumptions about privacy align with the regulatory definitions and with the technical processes that occur in practice. We follow Westin (Westin, 1967) in treating privacy not as secrecy but as a claim to control over how information about oneself is communicated, a definition that locates the failure mode of consent not in withheld information but in the gap between perceived and actual control. Where the three perspectives (user, regulatory, technical) diverge, that gap weakens meaningful consent and limits users' capacity to reason about how their data is collected, processed, and reused. Existing research has documented each perspective on its own terms. As Milne and Culnan (Milne and Culnan, 2004), Bandara, Fernando and Akter (Bandara et al., 2019) and Oates et al. (Oates et al., 2018) show, users routinely misunderstand privacy controls and consent mechanisms; the privacy paradox is partly sustained by mechanisms such as psychological distance, perceived social contracts, and learned helplessness, as discussed by Bandara, Fernando and Akter (Bandara et al., 2019) and Culnan and Armstrong (Culnan and Armstrong, 1999); and technical and legal analyses have proceeded largely in parallel, as Schaar (Schaar, 2010) argues. The three are seldom brought together through concrete demonstrations of the gap between perceived and actual privacy, as Barth, Ionita and Hartel (Barth et al., 2022) and Guo, Rodolitz and Birrell (Guo et al., 2020) note.

The contribution is a visual language for comparing the three perspectives on the same template, built from four-axis data flow cards (What, From, How/Why, Time) that mirror parts of what controllers must already document under GDPR Articles 13, 14 and 30, bidirectional arrows that keep the return path of data visible, an onion topology where distance from the centre encodes opacity from the user's point of view, and a three-case set (A: Cookies, B: Login, C: Chat) anchored in rulings from the Danish Data Protection Agency (Datatilsynet, 2023b, 2022, 2021b, 2023a, 2021a). The proof of concept is validated as a research instrument across seven short intercept interviews with non-technical participants (P1-P7) on a single news-site mock-up (`example-news.com`); the technical reading per case is documentary, drawn from the regulator's own evidence rather than from independent technical measurement.

The sessions surface a consistent pattern: participants map data flows to the primary service but rarely extend the chain beyond it. Where the regulatory and technical pillars document onward transfers to sub-processors, advertising networks, and data brokers, the user pillar stops at the first step. This stopping point does not reflect ignorance in the ordinary sense; participants generally acknowledged that further links in the chain existed. Rather, it reflects what one participant named *lykkelig uvidenhed* (blissful ignorance) a deliberate strategy of non-engagement that participants articulate and sometimes label as naive while still adopting it. The privacy paradox emerges not as a simple inconsistency between stated values and behaviour, but as a considered adaptation to a system perceived as too large and too opaque to act on.

The thesis proceeds as follows. The literature review situates the study in existing research on privacy perception, consent mechanisms, and data-flow visualisation. The methodology section presents the visual language and the research approach. The results section reports the findings from the seven sessions across the three cases and the cross-case patterns they reveal. The discussion section interprets those findings in relation to the three pillars. The limitations section reflects on the constraints of the proof of concept, and the implications section draws out what the findings suggest for consent design and future research. The thesis closes with a conclusion.

3 Literature Review

Privacy is a concept that resists a simple or stable definition. It shifts depending on context, culture, and technology, and the academic literature reflects this instability rather than resolving it. For the purposes of this project, three dimensions of privacy are particularly relevant: how non-technical users understand and experience it, how it is defined by regulation, and how it actually plays out in technical practice. The following sections trace what existing research tells us about each of these dimensions, and identify the gap that motivates the present study.

3.1 Privacy as a Contested Concept

The difficulty in establishing a definition of privacy is not incidental, but structural. Privacy functions at once as a personal value, a legal right, and a social practice, and these do not always point in the same direction, as Bennett (Bennett, 2011) argues. A user may feel a strong sense of privacy violation without any legal threshold being crossed, while a data practice that is perfectly compliant on paper may still involve processes that most people would find alarming if they were made visible. This gap between the felt, the legal, and the technical is exactly what this project sets out to examine.

A widely cited attempt to ground privacy in a working definition comes from Alan Westin, who in 1967 described privacy as *"the claim of individuals...to determine for themselves when, how and to what extent information about them is communicated to others"* (Westin, 1967). What makes this definition useful is that it places control, rather than secrecy, at the centre

of privacy. Privacy, in Westin's account, is not about hiding information but about having meaningful autonomy over how it flows. This distinction matters in a digital context, where users often share information willingly but have no realistic understanding of what happens to it afterwards, and no practical ability to change that even if they did.
Westin further proposed four states through which individuals experience and seek to manage privacy in practice; solitude, the state of being free from observation by others; intimacy, the ability to share within a small and trusted group; anonymity, the freedom to act in public without being identified or tracked; and reserve, a psychological boundary that allows individuals to withhold information without having to justify doing so (Westin, 1967). These four states are not equally available in digital environments. Solitude is undermined by background data collection that operates regardless of whether a user is actively engaging with a service. Anonymity is undermined by tracking technologies that link behaviour across sessions and platforms. Intimacy is complicated by the fact that private communications are stored on third-party servers and subject to their data practices. Reserve is perhaps the most directly relevant here: when a user is confronted with a cookie banner or a login screen, the default assumption is that they will engage and explain, not that they can simply decline without cost.

Understanding privacy in this way also requires taking context seriously. Nissenbaum's (Nissenbaum, 2010) concept of contextual integrity holds that privacy is not about keeping information secret, but about whether information flows in ways that match the norms of the context in which it was originally shared. When a news site passes a reader's browsing behaviour to an advertising broker, this may not be illegal under all circumstances, but it likely violates the contextual expectations under which the reader understood themselves to be engaging. The reader came to read an article, not to enter into a data-sharing arrangement with a network of third parties they have never heard of and cannot name. Mapped onto Westin's account, this kind of invisible data transfer undermines anonymity and reserve simultaneously: the user is identified and tracked without meaningful awareness, and has no practical way to withhold that information flow without abandoning the service entirely.
Privacy is also not experienced uniformly across interactions. The same user who thinks carefully about what they share on a sensitive platform may click through a cookie banner without a second thought on a news site, treating the two situations as categorically different even though the underlying data flows may be equally extensive.

3.2 The Privacy Paradox and User Behaviour

A recurring theme in the literature is the gap between how much people say they care about privacy and how little that concern seems to affect their actual behaviour. This phenomenon is commonly referred to as the privacy paradox. Culnan and Armstrong (Culnan and Armstrong, 1999) showed early on that users who perceive fair information practices are more willing to share personal data, which suggests that the subjective experience of how data is handled matters independently of how it is actually handled. Perception shapes behaviour at least as much as reality does, and in some cases more so.
This finding connects directly to Westin's account of privacy as control. If users perceive that they have a degree of meaningful control over their information, they become more willing to share it, regardless of whether that perception is accurate. The architecture of consent

mechanisms, including cookie banners, privacy settings, and login choices, therefore functions less as a genuine means of control and more as a signal of trust. The feeling of having been given a choice matters more to user behaviour than the actual content of that choice.
Bandara, Fernando and Akter (Bandara et al., 2019) offer one of the more thorough treatments of the privacy paradox. Through a qualitative study of online shoppers, they identify three mechanisms that help explain why people continue to act against their own stated privacy interests. The first is psychological distance: privacy harms tend to feel remote in time and consequence, so users prioritise the immediate reward of access over an abstract future risk. The second is perceived social contracts: most users operate with an implicit expectation that their data will be handled ethically, even when no such guarantee exists. The third, and most significant for this project, is learned helplessness: a state in which repeated exposure to vague and uncontrollable data practices leads users to stop engaging with privacy management altogether, not because they no longer care, but because they have come to believe that caring makes no practical difference.

3.3 Privacy Visualisation and Communication

If the privacy paradox is partly sustained by the invisibility of data flows to non-technical users, then a natural response is to make those flows more visible. Several attempts have been made to do exactly this, and they reveal both the potential and the real limits of visual approaches to privacy communication.

Barth, Ionita and Hartel (Barth et al., 2022) provide a comprehensive overview of 41 privacy visualisations and 14 Privacy by Design guidelines in a systematic survey. Their central finding is that most existing tools only show the relationship between the user and the first service they interact with, and stop there. What happens after that, where the data goes next and who ends up receiving it, is largely left out. This matters because the parts of the data chain that are furthest from the user are also the least visible and the hardest to opt out of. A tool that only shows the first step may give users a false sense of understanding, making them feel they have a grasp of what is happening to their data when in reality the more consequential parts of the chain have simply not been shown to them.

A concrete example of a visual tool in this space is the Privacy Rating developed by Barth et al. (Barth et al., 2021), which presents the data-handling practices of online services through twelve privacy attributes grouped into four clusters: collection, sharing, control, and security. Their evaluation found that such visualisations can genuinely improve user engagement and awareness, and that users respond positively to structured and clear representations of what would otherwise be buried in dense policy documents. The study also showed, however, that the design of the tool actively shapes what users pay attention to. This is a finding with direct methodological relevance to the present study, since it means that any visual language will inevitably influence the responses it produces to some degree.

Guo, Rodolitz and Birrell (Guo et al., 2020) present Poli-see, an interactive graphical representation of privacy policies, evaluated against both annotated text and plain text alternatives. Participants using Poli-see found it more engaging and more enjoyable to interact with, and were more likely to say they would consult such a representation in practice. This

is an encouraging result for the broader project of making privacy information accessible to non-technical users. The study also surfaced important challenges: users unfamiliar with the graphical format faced a steep learning curve, and several found the number of links and connections overwhelming. There was also confusion around vague or non-specific policy rules, where the tool struggled to represent the fact that many real-world policies apply broadly across all data types rather than specifying clear flows between specific categories. Finally, the tool faced a practical scalability problem, since generating a Poli-see representation requires a domain expert to encode each privacy policy manually, which limits its potential for widespread use.

Finally, Oates et al. (Oates et al., 2018) examine how both experts and non-experts represent privacy visually through illustrations, and find systematic differences in the mental models each group draws on. Non-experts tend to reach for physical and spatial metaphors such as walls, locks, and envelopes, while experts are more likely to represent privacy as a network or a flow of data between entities. These differences are significant for tool design, because they suggest that a visual language aimed at non-technical participants should be built around spatial and relational structures that feel intuitive, rather than around abstract system diagrams that require technical skills to interpret.

3.4 The Gap This Project Addresses

Taken together, the literature points to three things. First, as Bandara, Fernando and Akter (Bandara et al., 2019), Culnan and Armstrong (Culnan and Armstrong, 1999) and Barth, Ionita and Hartel (Barth et al., 2022) show, users routinely misunderstand the scope of data collection and sharing in their everyday digital interactions, and in our reading, this misunderstanding tends toward underestimation. This misunderstanding is not simply a matter of lacking information but is shaped by psychological mechanisms and design choices that actively make informed engagement difficult. Second, the visualisation tools developed so far, surveyed by Barth, Ionita and Hartel (Barth et al., 2022), Guo, Rodolitz and Birrell (Guo et al., 2020) and Barth et al. (Barth et al., 2021), approach this problem primarily from the perspective of the service provider, rendering a policy or a data practice readable after the fact rather than eliciting the user's own mental model before any ground truth is introduced. Third, the three perspectives that together shape what privacy actually is in any given situation (the technical reality, the regulatory definition, and the user's own mental model) are rarely brought into direct comparison using a shared vocabulary, as Barth, Ionita and Hartel (Barth et al., 2022) note. This project addresses that gap by developing a visual language that places all three perspectives side by side on the same template, making the divergence between them concrete and comparable.

4 Methodology

The research question asks how far non-technical participants' perceptions of how their personal data is processed align with the regulatory and technical reality. To answer it, the methodology puts all three perspectives into the same vocabulary so they can be compared on common terms. The rest of this section walks through the design in four steps: the visual language itself, how the three interview cases were constructed from rulings by the Danish Data Protection Agency (DDPA), how the interviews were run, and how the resulting material is analyzed.
Three ideas shape the design. The first is that privacy is about control over information flow, not secrecy, so the elicitation must surface what the participant thinks is happening to their data rather than whether they want that data kept hidden. The second is that non-technical users reason about privacy in spatial and physical terms, so the language uses data flow cards and arrows with prefilled options rather than a more complex diagram. The third is that participants who disengage from privacy questions tend to do so for predictable reasons (psychological distance, trust in a perceived social contract, and learned helplessness), and the coding scheme leaves room to label each route as it appears.
The empirical core is seven short (qualitative semi structured) in-person interviews (P1 to P7) with non-technical participants recruited on the campus of the University of Copenhagen, organised around three everyday personal-computer interactions: browsing (cookies), account login, and messaging (chat).

4.1 Development of the Visual Language

If we just ask a participant where their data goes, the answer comes back in a single mix of regulatory, technical and everyday language, with no way to separate the three. We need a structure that pulls them apart so they can be compared. The visual language aims to fix this by describing every recipient of data along the same four axes (What, From, How/Why, Time) and connecting recipients with bidirectional arrows. The four axes mirror some of what controllers must already document under GDPR (categories of data, recipients, purposes, retention), so the participant's mental model and the controller's lawful record sit on the same template. That is what makes the comparison between perceived, regulated, and technical privacy non-arbitrary.

Three existing systems inspired the design. The card-and-arrow grammar follows BlueJ, the beginner Java IDE used in our introductory programming course, and the related UML convention of boxes-with-content connected by arrows; the style is aimed at a beginner with no prior technical knowledge, and matches Oates et al.'s observation that non-experts represent privacy as spatial relationships between named entities rather than as network diagrams (Oates et al., 2018). The layered topology responds to Barth, Ionita and Hartel's finding that existing privacy visualisations cover the first step of a data flow far more thoroughly than the onward steps to third parties (Barth et al., 2022): the onion's layers put the under-covered steps on the same page as the first one. The icons in the four panels are taken from the Swiss Privacy Icons by the Privacy Icons Association (Privacy Icons Association, nd), an existing disclosure vocabulary used here under permission granted for academic purposes. The How/Why panel adds two standard Unicode glyphs (a gear for service operation, a balance

for legal obligation), and the Time panel uses four custom hourglass icons we generated using AI to mark the retention buckets (now, short, long, indefinite).
The conceptual model is an onion of recipients (Figure 1). At the centre sits the example case. Connected to the Core layer for the first party. Outside that sits the 2nd layer for disclosed third parties (entities reachable from the Core in one step because they appear in the visible interface, such as the SSO provider behind "Log in with Facebook", the chat-widget vendor, or partners listed in the cookie banner). Outside that sits the 3rd layer for undisclosed onward recipients (sub-processors, ad-tech partners, and brokers, visible only in the privacy policy or in real network traffic). Bidirectional arrows are deliberate: a single-headed arrow would suggest data only flows outward.

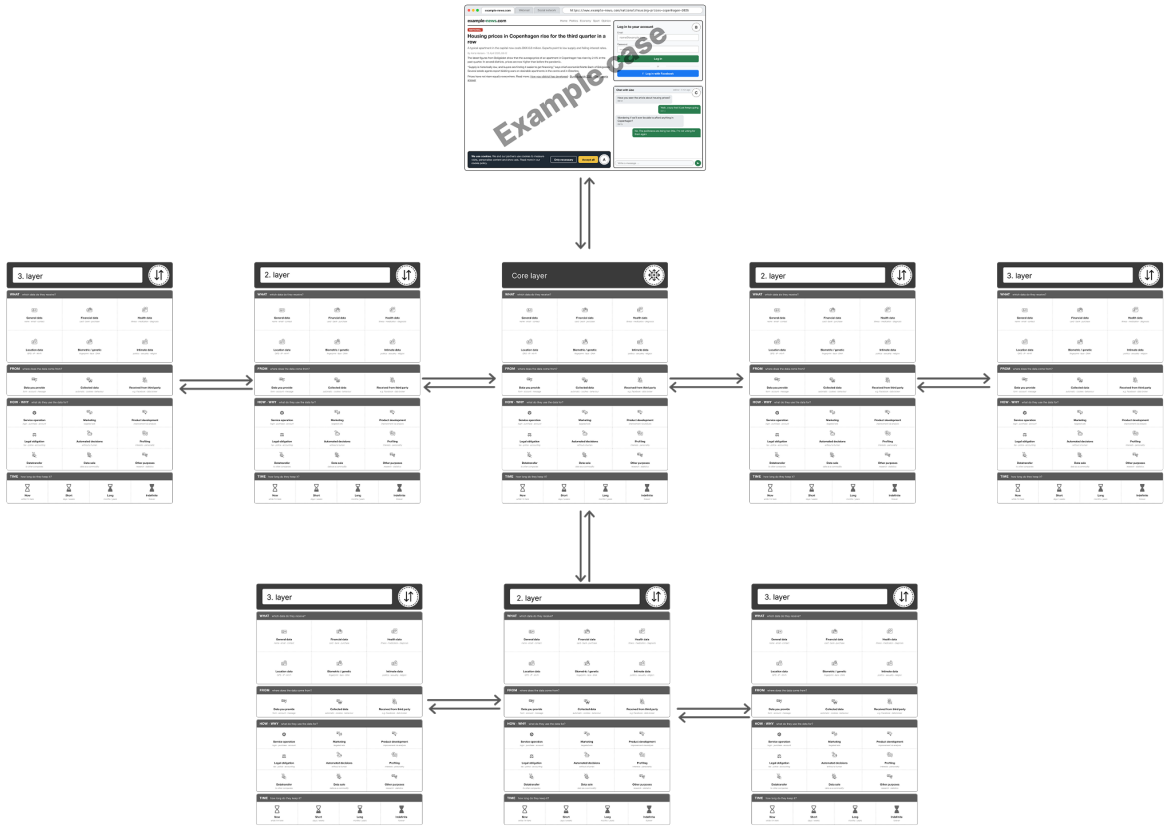


Figure 1: The framework's intended layout: data flows outward through the chain of layers. Every card uses the same four panels (What, From, How/Why, Time); every link is a bidirectional arrow.

4.1.1 What was intentionally cut

Several earlier ideas were dropped to avoid steering placements or forcing emotional commitments before the participant had finished thinking: coloured arrows for trust, pre-printed concentric layer rings on the context window, sticker tile kits, single-headed arrows, and a separate Where category nested over What. We also chose not to hand out a take-home debrief brochure after the interview; an earlier version of the kit included one with the DDPA

rulings for each case accordingly, but distributing DDPA rulings could have positioned us as offering legal guidance, which falls outside the competence of a BA-level study.

4.2 Case Construction from DDPA Rulings

The three cases anchor the regulatory and technical pillars. Each case starts from DDPA ruling, reduces it to a recognisable everyday scenario, and constructs structured context from the cases, whose answer marks with the letter of the current case context (A/b/C) to cover the assumption, the lawful baseline, and the technical truth, so that the mismatch is observable without telling the participant which answer is which. The scenarios were anonymised in the kit by creating a fictional website `example-news.com` to serve as basis for all three cases. The full case window can be seen in Appendix 6.

4.2.1 Case A: Cookies

The participant is told they visit `example-news.com` to read an article and click "Accept all" on the cookie banner. The case is anchored on the DDPA ruling against Texas Andreas Petersen (Datatilsynet, 2023b), in which tracking pixels and cookies from Google and Meta fired on the site and transferred personal data without valid consent under GDPR Article 6. The technical reality is that on click the browser sends the URL, IP address, browser fingerprint and a unique cookie identifier to each tracker, joining the visit to whatever Google and Meta already know about the participant.

4.2.2 Case B: Login

The login component shows two paths: ordinary email and password, or "Log ind med Facebook". The case combines two regulatory anchors. The password path is anchored on the DDPA rulings against Salling Group (Datatilsynet, 2022) and T. Hansen (Datatilsynet, 2021b), where customer passwords were stored in plaintext in violation of GDPR Article 32(1). The Facebook-login path is anchored on the DDPA ruling against Boligportal (Datatilsynet, 2023a), where Facebook Business Tools transferred user data to Meta beyond what users were aware of. The technical reality is that good practice stores passwords as hashes (one-way and not reversible). A page containing a Facebook login button triggers an HTTP request to Meta on load, carrying the URL, IP address and any existing Meta cookies before the participant has clicked anything or agreed to any data sharing.

4.2.3 Case C: Chat

The participant is told they use the news site's built-in chat with a friend, and the conversation drifts into political opinion. The case is anchored on the DDPA ruling against Dating.dk (Datatilsynet, 2021a), where user-to-user messages on a website were stored on the platform's servers and processed by automated keyword scanning without the explicit consent required for special-category data under GDPR Article 9. The technical reality is that browser-based chat is almost never end-to-end encrypted: TLS protects the message in transit, but the server stores it, and could potentially be read by the platform's IT staff.

4.3 Qualitative Interview Design

The interviews were designed as short semi-structured intercept sessions, a format chosen because it fits the proof-of-concept scope of the study and allows us to reach non-technical participants in a natural setting. Seven sessions were run in person on the University of Copenhagen campus.

Candidates were screened on the spot against mainly these criteria: 18 or older, regular laptop or PC use, and no prior study or professional work in computer science, IT security, or GDPR or compliance. Candidates with a background in digital marketing or analytics were considered on a case by case basis, since familiarity with advertising tools does not necessarily imply knowledge of the underlying data infrastructure. A judgement-based check for generally tech-savvy behaviour was also applied.

The semi-structured format means each session followed the same structure and covered the same three cases, but the moderators could follow up on interesting responses and let silences sit. The visual language did most of the structuring work: because every participant worked with the same cards and the same four axes, the sessions are comparable without needing a rigid script. Before any material was touched, three norms were named aloud: this is not a test, there are no right or wrong answers, and only the participant's own perception matters.

The interview kit consists of a context window showing `example-news.com` with three anchor points (**A** cookies, **B** login, **C** chat) and a set of data flow cards. The Core card is pre-named `example-news.com` and placed on the table at the start of each session, already connected to the context window by one bidirectional arrow. The remaining cards are blank-name templates that the participant fills in. For each case the moderator read the scenario aloud, the participant marked the relevant panels and placed any further cards they thought were relevant, and a photo was taken at the end of each case to record the layout.

4.4 Analytical Framework

The analysis compares what participants thought was happening with what could actually occur, based on three readings for each case. The first is the user reading: which data flow cards the participant placed, which layer they placed them on, and which panels they marked. The second is the regulatory reading, drawn from the relevant DDPA decisions (Datatilsynet, 2021a,b, 2022, 2023a,b). The third is the technical reading, which draws on the technical descriptions in the DDPA case files combined with our own background as software development students and a comparison against existing real-world sites to ground the technical claims in practice. Where the three readings diverge, that divergence is the mismatch the analysis is interested in. The interview transcripts were read alongside the card photos to find quotes that could support and expand on what the cards showed, giving the data more context where participants explained their choices or reflected on what they did not know.

5 Results and Analysis

Seven participants took part in intercept sessions of approximately 15 to 25 minutes each. All were recruited at the University of Copenhagen campus, were Danish speaking, and fell primarily within the 18 to 29 age range, with two exceptions. The sample is thus small, demographically narrow, and skewed toward students.

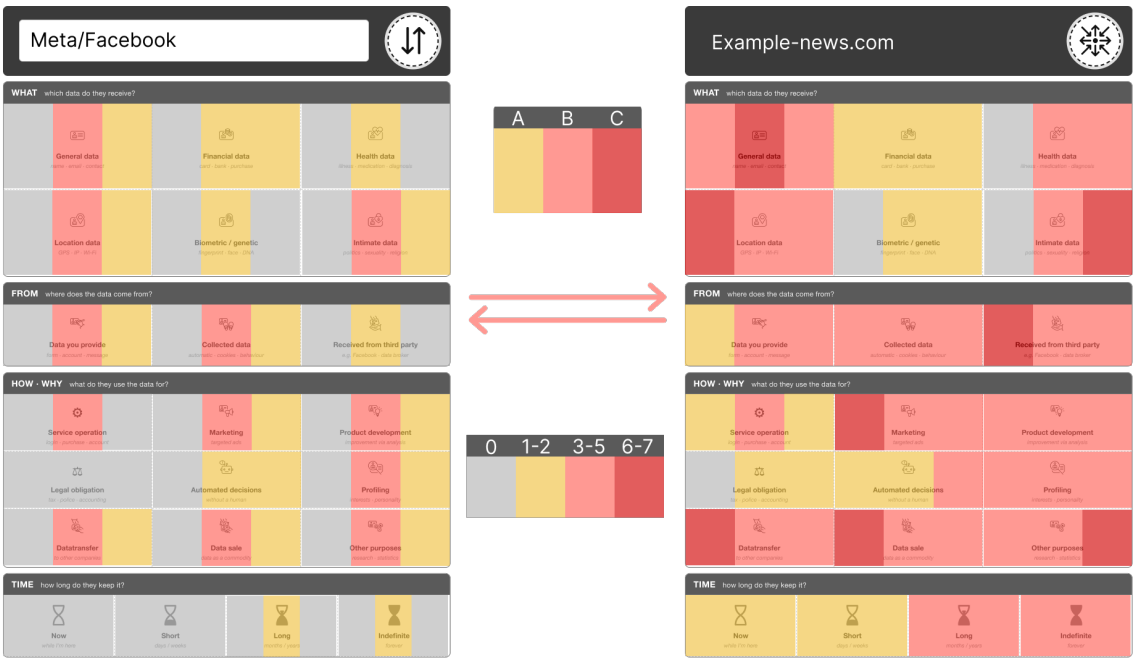


Figure 2: Aggregated recipient card responses across P1-P7 for all three cases. Left card: Meta/Facebook (N = 4, participants who extended the chain to a second layer). Right card: example-news.com (N = 7, all participants). Each icon cell is split into three vertical bands: A (Cookies), B (Login), C (Chat). Colour encodes how many participants marked that option: grey = 0, yellow = 1-2, pink = 3-5, red = 6-7.

Figure 2 aggregates recipient card responses across P1-P7 onto two cards. The example-news.com card on the right was filled out by all seven participants in every case. The Meta/Facebook card on the left was placed by only four participants, and first connected with the login case. Each icon cell on both cards is split into three vertical bands corresponding to Case A (Cookies), Case B (Login), and Case C (Chat), with colour encoding how many participants marked that option: grey means zero, yellow means one to two, pink means three to five, and red means six to seven. Because the two cards have different contextual weight. Pink on the Meta card means at least three out of the four added it, while pink on the example-news card means roughly half of the participants marked those fields. This pattern of stopping at the first step is a central finding of the study. The data chain that the regulatory and technical pillars document, involving sub-processors, advertising networks, and onward data brokers, is simply not present in the user pillar. This is not because participants believed it did not exist, but because they did not extend their mapping that far, and in several cases explicitly said they preferred not to think about it. It is also worth acknowledging that the visual language and the cases themselves may not have been explicit enough to prompt

participants to reach further along the chain.

5.1 Case Analysis

5.1.1 Case A: Cookies

The most striking feature of the Cookies case in the heatmap is the combination of a strong signal at the first step and a complete absence of signal at the second. On the example-news card, the A column lights up red on Location data, Received from third party, Marketing, Data transfer, and Data sale. Participants perceived cookies as a significant data-extraction event and were willing to commit to strong claims about what the news site itself receives and does with the data. But the Meta/Facebook card is entirely grey in the A column across every single panel. Not one of the four participants who added a Meta card at all placed it in the chain for the Cookies case. The data flow documented by the Datatilsynet decision against Texas Andreas Petersen (Datatilsynet, 2023b), where a Meta pixel fires on a site and transfers data to Meta without valid consent, is invisible to the user pillar at the point of card placement. The chain is cut before Meta is even on the table.

This creates a notable internal contradiction in the data. Most participants marked both Data transfer and Received from third party as red on the example-news card for the Cookies case, meaning they believed data was being transferred to third parties and that the site was receiving data from third parties. Yet none of them placed a second card to represent who those third parties actually are. Participants were willing to acknowledge that a third-party relationship existed in the abstract, but did not translate that acknowledgement into a concrete recipient on the table in front of them. This gap between recognising that something is happening and being able to name what or who is central to what the framework is designed to surface. The knowledge is there at a general level but it does not extend to the specific actors that make the data flow real. In Westin's terms, you cannot exercise meaningful control over a flow you can sense but cannot name (Westin, 1967).

Inspecting real Danish news sites such as DR.dk or EB.dk after accepting cookies shows primarily first-party activity: cookie consent management, session identifiers, and login state checks, with limited third-party tracking visible at the network level. The more extensive third-party flows documented in the regulatory cases tend to occur on smaller or commercially oriented sites where advertising revenue is the primary business model, which makes the Texas Andreas Petersen case a more accurate anchor for the regulatory pillar than the major public broadcasters. What this illustrates is that the sites participants are most familiar with may give them a relatively contained picture of what cookie acceptance involves, which could itself contribute to why the mental model of the data chain stays shallow. A site like DR.dk, which several participants named as a trusted source, does not produce the kind of visible third-party network activity that would make the onward chain feel real or immediate.
The TIME axis on the example-news card shows pink on Long and Indefinite for Case A, which is one of the more accurate readings in the dataset since cookies do in many cases persist for extended periods. Legal obligation is grey across all three cases on both cards, which suggests that the regulatory dimension of data collection is simply not part of how participants naturally think about these interactions. The framework made the category

available but participants did not reach for it, even though cookie consent itself is a legal obligation under GDPR Article 6, meaning that the very act of presenting a cookie banner is required by law, not a courtesy.
The dominant behavioural pattern was not careful consideration but habitual acceptance driven by a preference for ease of use. P5 put it directly:

> *"Jeg må indrømme, jeg laver nok tit den dovne bare trykker [ja]. Men det er altid på en måde bliver man lidt straffet, fordi så er det sådan, at så kan man ikke se et kort. Ellers kan man ikke interagere med nogle ting."*
> "I have to admit, I often do the lazy thing of just clicking [yes]. But you're always sort of punished, because then you can't see a map. Otherwise you can't interact with anything." (P5, 3:36-3:48)

The consent action here is not meaningfully informed. It is a response to the design of the interface. The mention of a map is telling: what P5 most likely encountered was a page embedding a third-party map service, where declining non-essential cookies blocks the map widget because it relies on cookies from the map provider. In practice the full site remains accessible, but the third-party component does not load. The perception that the entire page breaks is itself a product of how the consent interaction is designed, and it produces exactly the coercive dynamic that the GDPR standard of freely given, specific, and informed consent under Article 4(11) (European Parliament and Council of the European Union, 2016) is meant to prevent. The participant knows the click is uninformed and makes it anyway, not because they cannot decline, but because the interface makes declining feel like a punishment.

5.1.2 Case B: Login

The Login case produced the only second-layer extension in the dataset. The four participants who added a Meta card did so primarily here, and once they did, they marked it consistently across multiple axes. The B column on the Meta card shows pink on General data, Location,intimate, Data you provide, Collected data, Service operation, Marketing, Profiling, product development, data transfer, data sale and Other purposes, meaning most who added the card committed to those categories for the login case. This is the most information-rich part of the Meta card in the heatmap, and in all cases it was only reached after some mild to heavy prompting during the session.
The example-news B column has one red cell on General data and otherwise shows pink across most of How/Why and From, making the Login case the least dramatically lit at the first step even though it is the case that motivated the only second-layer extension. There is an interesting asymmetry here: login prompted participants to think about a second recipient but did not produce the strongest first-step signal on the primary card. The act of logging in with Facebook made the Meta connection feel relevant in a way that cookie tracking on the same site did not, even though both can involve data flows to potential third parties such as Meta.[1]

[1]The login screen in our mock case explicitly presented a Facebook login button, which was a deliberate design choice to give participants at least one named connection to work with. Even with that prompt built directly into the case, only four of the seven participants extended the chain to a Meta card, and in those cases additional nudging from the moderators was still required.

Five out of seven participants said they would be inclined to use or prefer Facebook login, primarily for convenience. The reasoning was consistent: creating a new account takes time, and Facebook login is faster. Several participants acknowledged without prompting that this felt like the less private option but chose it anyway. This is a direct instance of the privacy paradox as described by Bandara, Fernando and Akter (Bandara et al., 2019): the user recognises the trade-off and makes it anyway because the immediate benefit of convenience outweighs the abstract cost of reduced privacy.

It is worth noting, as our own observation rather than an established finding, that the choice between Facebook login and email registration is not straightforwardly one sided from a privacy perspective. Facebook login concentrates the data relationship with a single party that already holds extensive profile information and can link the new login to existing behavioural data across platforms. Email registration creates a new and separate relationship with the service, which may carry its own data practices but does not automatically feed into an existing cross-platform profile. Neither option is clearly better from a privacy standpoint, but Facebook login tends to be perceived as the obviously worse choice, while the risks of email registration are less visible precisely because the service is less well known.
The Datatilsynet cases anchor both sides of this picture. The Boligportal case (Datatilsynet, 2023a) establishes that using Facebook Business Tools for login involves data transfers to Meta that users are rarely aware of and that are not always adequately disclosed, which supports the concern around Facebook login. The Salling Group case (Datatilsynet, 2022) points in the other direction: Salling was criticised for storing user passwords in plain text in their internal logs, a basic development error that exposed account credentials without any third-party involvement. Taken together, the two cases illustrate that neither route is straightforwardly safe. Facebook login hands data to a large platform with its own extensive data practices. Email registration places trust in the service itself, which may handle that data poorly regardless of intent. The asymmetry in how participants perceived the two options, Facebook as the obvious privacy risk and email registration as the responsible alternative, does not fully hold up against the regulatory record. This gap between perceived and actual risk is itself a finding about how the user pillar assigns safety.

5.1.3 Case C: Chat

The Chat case produced two red cells on the example-news card: Intimate data and Other purposes. The Intimate data finding is notable because it suggests participants applied a stronger sensitivity frame to chat content than to cookies or login. People understand that messages feel personal in a way that browsing behaviour does not, and the visual language gave them a vocabulary to commit that intuition to the card. However, the regulatory category that covers this kind of content, the special categories of personal data under GDPR Article 9, was not reached in conversation by any participant. The gap between marking Intimate data on the card and articulating the Article 9 framework is the clearest user-to-regulatory mismatch in the dataset for this case.
The Meta card is mostly grey in the C column, with only yellow on Long and Indefinite in the TIME panel, marked by one of the four participants who had a Meta card at all. Participants did not strongly perceive Meta as a recipient of chat data, even though the technical architecture of a real chat feature would involve server-side storage subject to the same third-party

sharing logic as the other cases. The chain for the Chat case was contained to the first step in the same way as for Cookies, which reinforces the finding that the stopping point is not case-specific but reflects a general tendency to end the mental model at the primary service.

Participants expressed a general sense that chat messages were being read or processed in some way, but the mechanism was vague and the actors were unnamed. The dominant assumption was that the website itself was reading or storing the messages, rather than any third party. Several connected the case loosely to experiences of seeing targeted advertisements that seemed to reflect recent conversations, which points to a folk theory of surveillance that captures something real about how behavioural data is used but is imprecise about which parties are involved.

This assumption, that content privacy is primarily about what the service itself does with messages, mirrors a finding from Griggio, Nelson, Sramek and Askarov (Griggio et al., 2026), who show that users tend to focus on content and visible recipients rather than on the metadata layer surrounding a conversation. Their study finds that users broadly underestimate how much can be inferred simply from who communicates with whom, when, and how often, without any access to message content at all. While their study concerns messaging platforms rather than embedded chat features, the underlying pattern is the same: the assumption that if the content feels private, the data flow is private too.

The broader debate around chat control is worth briefly reflecting on here, not because our participants raised it directly, but because it maps onto the same tension their responses point toward. Several participants expressed conditional acceptance of the idea that authorities might access communications in serious circumstances, which echoes the optimist persona identified by Griggio et al. (Griggio et al., 2026): users who support privacy in principle but endorse exceptions when safety or law enforcement is invoked. The EU-level proposals to require scanning of encrypted communications provoke exactly this tension. What our participants showed is that this conditional attitude applies not just to state surveillance but to commercial processing as well: the chat feature on a news site felt acceptable to engage with precisely because the participants assumed its reach was limited to the site itself, a trust that the regulatory and technical pillars do not straightforwardly support.

5.2 Cross-Case Patterns

5.2.1 Blissful Ignorance: Not-Knowing as the Default Mode

Across the seven sessions, the most consistent pattern at the user pillar was not wrong knowledge but a conscious, articulated absence of it. Participants did not present themselves as misinformed; they presented themselves as deliberately not engaging with the question of where their data goes. This posture maps directly onto the learned helplessness mechanism identified by Bandara, Fernando and Akter (Bandara et al., 2019): users do not disengage from privacy because they do not care, but because repeated exposure to opaque and uncontrollable systems has taught them that caring produces anxiety without producing any practical ability to change the outcome. The mismatches surfaced by the framework sit on top of this stance, and the quotes below name it directly.

The clearest articulation came from P1 in the closing reflection of the session. After remarking that the chain probably had "mange flere grene end dem jeg lige har lagt ud og mange flere lag til" ("many more branches than the ones I just laid out, and many more layers"), the participant continued:

> *"Og at sådan jeg synes bare det lyder sådan lidt skræmmende at skulle gå for meget ind i, hvad det mon egentlig handler om. Og så tror jeg bare, at jeg tænker okay, så lever jeg bare i lykkelig uvidenhed på det punkt."*
> "It just sounds a bit scary to have to go too deep into what it might really involve. And then I just think, okay, then I'll just live in blissful ignorance on that point." (P1, 18:43-18:52)

The phrase *lykkelig uvidenhed* is the participant's own and was offered as a description of an active strategy, not a confession of a gap. In Westin's terms, this is the collapse of reserve as a functional state (Westin, 1967). Reserve, the ability to withhold information without having to justify doing so, presupposes that the individual has a realistic sense of what withholding would mean and what it would cost. When the chain is perceived as too large and too opaque to map, reserve stops being a choice the user can make and becomes instead a background anxiety that is easier to suppress than to act on. P1 named the mechanism explicitly earlier in the same session:

> *"Jeg tør ikke helt at undersøge, hvad det egentlig gør, fordi så tror jeg, jeg bliver for bange for at finde ud af, hvad det er jeg egentlig sidder og klikker ja til."*
> "I don't quite dare to look into what it actually does, because then I think I'd get too scared to find out what it is I'm actually sitting and clicking yes to." (P1, 17:22-17:34)

This is the consent gap the project is arguing for, in the participant's own words. The thing the participant is clicking yes to is, by their own account, something they have actively chosen not to investigate, on grounds that investigating it would produce fear without producing control. P4 articulated the same gap as a matter of behaviour rather than affect:

> *"Nogle gange så tror jeg bare, at [...] man gør en masse, som man måske ikke ville gøre, hvis man havde lidt mere viden om det."*
> "Sometimes I just think, you do a lot of things that you maybe wouldn't do, if you had a bit more knowledge about it." (P4, 18:25-18:35)

Together, P1 and P4 give both halves of the same finding from two participants independently: the user pillar chooses not to engage with the chain past the first step, and suspects, when asked, that the choice is shaped by that very non-engagement. This is precisely the kind of context-dependence that Nissenbaum's framework of contextual integrity helps explain (Nissenbaum, 2010). In a news reading context, the relevant social norm is a reader-publisher relationship. The expectation is that visiting a site to read an article does not constitute entering into a data-sharing arrangement with a network of unknown third parties. Because that norm is violated invisibly, without any signal that would alert the user to the mismatch, the user has no practical basis on which to adjust their behaviour. The blissful ignorance P1 describes is not a failure of attention but a rational response to a system that has consistently

withheld the information that would make attention useful. P7 made the contextual nature of the stance explicit: *"jeg tænker ikke så meget over det. Det er kun lige, fordi det er den situation"* ("I don't think much about it. It's only because of this situation right now") (P7, 17:32). The interview itself was framed as the unusual condition under which the participant had to think about a flow they otherwise do not, which is a direct expression of how context shapes what privacy concerns become more significant.

Where the technical chain was opaque, participants substituted brand trust as the proxy. This connects directly to the privacy paradox as described by Culnan and Armstrong (Culnan and Armstrong, 1999): users who perceive fair information practices become more willing to share data, regardless of whether that perception is accurate. The feeling of trustworthiness substitutes for actual knowledge of data practices. P4 named this substitution explicitly:

> *"Men når du siger DR, så kan det være jeg har lidt mere tillid til DR. Det er rigtigt. Naivt måske, men hvis nu hvor det var en anden side, som jeg havde mindre tillid til..."*
> "But when you say DR, maybe I have a bit more trust in DR. That's right. Naive perhaps, but if it was a different site, one I had less trust in..." (P4, 9:42-9:50)

The self-correcting "naivt måske" is doing analytical work here. The participant labels their own stance as naive in the same breath as adopting it, which is the same epistemic move as P1's lykkelig uvidenhed. Both participants are aware that their mental model is incomplete and choose to operate with it anyway, because the alternative, engaging seriously with the full chain, feels both frightening and futile. This is Westin's notion of control inverted: rather than determining when and how information about them is communicated, these participants have arrived at a position where they have quietly given up control of that determination and reframed it as a coping strategy (Westin, 1967).
P3 located the opacity in the system rather than the participant, which shifts the analysis from individual psychology to structural design:

> *"Det er umuligt at læse sig igennem en privatlivspolitik på en hjemmeside. Man kan godt klikke ind på den, og så kan man scrolle rundt, men man kommer aldrig i bund med den, fordi den er, den er rodet. Jeg ved ikke, om det er med vilje, men der er så meget tekst, at det er, det er en jungle."*
> "It's impossible to read your way through a privacy policy on a website. You can click into it and scroll around, but you never get to the bottom of it, because it's, it's a mess. I don't know if it's deliberate, but there's so much text, it's, it's a jungle." (P3, 14:53-15:13)

The suspicion that the opacity might be "med vilje" (deliberate) frames the privacy policy not as a failed communication but as an instrument designed to foreclose meaningful reading. This is consistent with what Barth, Ionita and Hartel document in their review of privacy disclosures: the onward chain past the first party is systematically under-covered in existing visualisations and policy texts, which means the incompleteness P3 experiences is not incidental but structural (Barth et al., 2022). P2 generalised the same point to the data ecosystem as a whole:

> *"Trods alt jeg har en oplevelse af, at det er så uoverskueligt alligevel, og man har så lidt kontrol over det... jeg ved alligevel ikke rigtigt, hvor alt min data ligger henne, og det, den bliver sikkert høstet et eller andet sted... den er så stor, at man ikke rigtigt kan tage stilling til den."*
> "After all, I have a feeling that it's so unmanageable anyway, and you have so little control over it... I don't really know where all my data is anyway, and it's probably being harvested somewhere... it's so big that you can't really take a position on it." (P2, 12:31-12:55)

The closing sentence is the explicit mechanism: the chain is not just opaque, it is too large for a position to be taken at all. In Westin's terms, the individual's claim to determine when and how information about them is communicated has become unworkable in practice, not because the legal right has been removed but because the scale and complexity of the system makes meaningful exercise of that right impossible for an ordinary user without specialist knowledge (Westin, 1967).

Read together, these quotations give the lived-experience reading of the first-step-only reach pattern. All seven participants articulated the same posture in their own words. The chain past the first step was named as opaque (*en jungle*, P3), as scary to engage with (*skræmmende*, P1), as something one is not aware of in everyday life (*man er ikke opmærksom på*, P4), as too large to take a position on (P2), as something one only thinks about under the unusual condition of being asked (P7), and as a click one makes because the alternative appears to break the page (P5). Lykkelig uvidenhed is not the absence of a model. It is a model: a deliberate first-step-only model, propped up by brand trust and contextual inattention, that participants can articulate and sometimes label as naive while still adopting it. The privacy paradox (Bandara et al., 2019) is visible here not as a simple inconsistency between stated values and behaviour, but as a considered adaptation to a system that has made informed engagement structurally difficult.

5.3 Data Credibility

The findings reported here should be read with several constraints in mind. The sample consists of seven participants, primarily students between the ages of 18 and 29, recruited through intercept on the University of Copenhagen campus, with two participants outside that age range. This is not a representative sample of Danish internet users or of any broader population, and no claim of prevalence is made on the basis of these sessions. The findings speak to the shape and character of the mismatches the framework surfaces, not to how common they are.

As the moderators of the sessions ourselves, we cannot fully rule out that our own conduct during the interviews shaped some of the responses we received. Some prompting occurred, particularly in the Facebook login case, and the checklist format of the recipient cards likely led participants to mark categories they would not have raised on their own. Both of these effects push the data toward over-representing perceived privacy concern, which means the findings should be treated as an upper-bound indication rather than an accurate measurement of what participants actually believe.

The sample is also linguistically and culturally homogeneous. All participants were Danish speaking, all sessions were conducted in Danish, and the regulatory anchors are drawn from

Danish Datatilsynet decisions and EU law. The findings are grounded in a specific national and regulatory context and may not transfer directly to other settings.

6 Discussion

6.1 Over-selection as a Sign of Not Knowing

A broad interpretive assumption running through much of the privacy literature, and one that shaped our own starting point, is that users underestimate the scope of data collection. Our results point toward a different picture. Participants did not leave fields blank when they were uncertain. They marked broadly, often committing to categories like Location data, Data sale and Received from third party across all three cases on the example-news card, not because they had a clear picture of what was happening, but precisely because they did not. When you do not know what a site collects, marking everything feels safer than marking nothing. The result looks like high privacy awareness on the heatmap, but it is better read as a different kind of not-knowing, one that expresses itself through over-commitment rather than avoidance.

This is consistent with what participants said in the sessions. Several acknowledged that they did not really know what was happening but suspected it was a lot. P2 described the data ecosystem as so large that you cannot really take a position on it. P3 called privacy policies a jungle. These are not the words of someone who has underestimated the situation. They are the words of someone who has given up trying to estimate it accurately and defaulted to assuming the worst. The visual language made that assumption visible by giving it a place to land, but the assumption was already there.

This connects to the learned helplessness mechanism identified by Bandara, Fernando and Akter (Bandara et al., 2019). Users who have repeatedly found privacy information overwhelming and inaccessible do not arrive at calm indifference. They arrive at a kind of resigned pessimism, a feeling that everything is probably being collected, that it is too complicated to understand, and that nothing they do will change it.

6.2 What the Visual Language Reveals and What It Does Not

The over-selection finding also raises a question about the visual language itself. The checklist format of the data flow cards made categories available for confirmation that participants would not have named spontaneously. When Location data and Data sale are printed on the card in front of you, marking them is easier than naming them from memory. We cannot rule out that the format amplified the over-selection tendency rather than simply recording it.

What makes the finding robust despite this concern is what participants did not mark. Nobody placed a Meta card for the Cookies case, even with the checklist format available and even though the Meta pixel firing without valid consent is precisely what the Texas Andreas Petersen decision (Datatilsynet, 2023b) documents. If the checklist format were simply priming broad confirmation, we would expect to see the same pattern on the Meta card as on the example-news card. We do not. The absence of Meta in the Cookies case is therefore a genuine finding: participants could articulate a pessimistic picture of what the primary service

does, but the specific onward transfer that makes that picture real in regulatory terms was not part of their model at all.
This is where the visual language contributes something the existing literature does not fully capture. By separating the first-step picture from the onward chain, it makes visible not just how much participants know but where their knowledge stops. The heatmap does not show users who underestimate privacy risk. It shows users who have a strong but inaccurate picture of the first step and no picture at all of what comes after it. That distinction matters for how we think about privacy communication and tool design going forward.

7 Limitations

The study is based on a small sample of seven participants, recruited through intercept on a single university campus, which restricts how representative the results can be considered. All participants were Danish-speaking, non-technical users encountered in one specific setting, and the three cases examined were built around one mock news site. This means the findings speak to the existence and shape of the mismatches the framework surfaces, not to their occurrence in any wider population.
The qualitative interview design introduces additional limitations around how participant responses were shaped. The interviews were informed by principles of semi-structured qualitative interviewing, with an emphasis on open-ended questions and non-directive probing. However, the intercept format and the time pressure of 15 to 25 minute sessions meant that full adherence to this ideal was not always possible, and some interviewer prompting occurred in practice. This should be taken into account when reading the interview material.
The checklist format of the data flow cards is a structural limitation. When categories are presented visually, participants are more likely to confirm them than to name them spontaneously, which means the cards likely produce an upper-bound estimate of perceived privacy concern rather than a representative one. The Facebook login case illustrates this directly: no participant placed a Meta card spontaneously, but all four who had one engaged with it extensively once introduced, suggesting the connection exists in their mental model but needs a guidance to surface. Findings should be read as directionally useful rather than precise.
Finally, the study focuses on participants' perceptions of privacy rather than their actual behaviour. What people say they think is happening to their data, and what they actually do when faced with a cookie banner or a login screen, are related but distinct. The interview format captures the former, not the latter, and this gap should be kept in mind when interpreting the findings.

8 Implications and Future Work

The study demonstrates that a structured visual language can surface the shape of a participant's mental model of personal data flows in a short session, and make the gap between that model and regulatory or technical reality directly comparable. The tendency of all seven participants to stop their mapping at the primary service and not extend to further recipients, reproduced across all three cases, would arguably not have been visible in a verbal interview alone: the onion topology gives the stopping point a location.
Privacy communication tends to assume that the problem is one of information access, that users would engage more meaningfully if policies were clearer or shorter. The findings point in a different direction: the gap is structural and psychological, shaped by learned helplessness and deliberate non-engagement, and closing it requires tools that meet users where their mental models actually stop rather than where policy assumes they start. Several refinements would strengthen a future iteration. The How/Why panel icons for closely related categories such as profiling and automated decision-making need clearer distinction. The Time buckets cover four retention levels (now, short, long, indefinite) but showed limited granularity across cases where longer-term retention was at stake; a scenario covering data retention after account deletion was researched but cut to keep sessions short and could be reinstated in a future iteration. A guided prompt inviting participants to extend the chain past their first placed recipient would directly address the first-step tendency rather than leaving extension as an open-ended option. A larger and more demographically varied sample recruited outside university campuses would test whether the same stopping point appears across different contexts.

The same grammar of four axes, bidirectional arrows, and onion topology also points toward two further uses that this study does not validate but that follow naturally from the design. A descriptive use would allow a designer, regulator, or journalist to fill in the data flow cards for an existing service and hand the result to a non-technical audience as a structured single-page summary of where data goes. A teaching use, aimed in particular at upper-secondary education, would add the one thing the research protocol deliberately withholds: feedback. A learner builds the onion for a given scenario, and on submission their build is compared against a known ground truth in the same vocabulary they used to construct it. The most important constraint for that tool follows directly from the findings: it should explicitly reward extending the chain past the Core, since that is the move participants most consistently failed to make.

A small interactive prototype was prepared alongside the printed kit as a concrete starting point for further development. The single-page web application `index.html` mirrors the printed kit on a pannable, zoomable canvas. It was developed with AI coding assistance and is intentionally limited: it reproduces the build interface only and is not part of the empirical contribution. The prototype is bundled with the thesis materials and is also available through the project's public GitHub (See appendix 2).

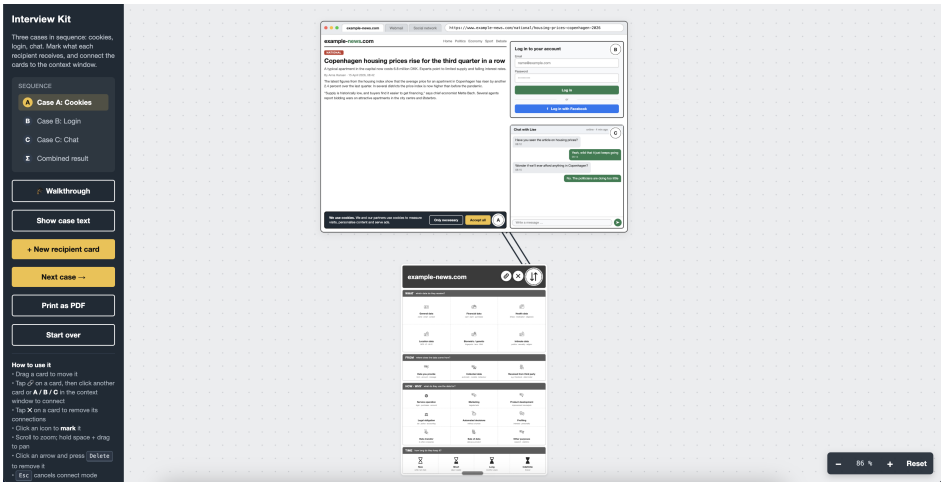


Figure 3: Screenshot of the reference prototype `index.html`: An app version of how we conducted the interviews using the visual language. This should be seen as a small AI-assisted artefact, not part of the empirical study.

9 Conclusion

This thesis set out to examine how non-technical users perceive the flow of their personal data in everyday digital interactions, and whether a structured visual language could help make the gap between that perception and regulatory and technical reality more concrete and comparable. Across seven intercept interviews built around three DDPA-anchored cases, what we can say is modest. The visual language seemed to work as a proof of concept, at least within the conditions of this study. Placing the user, regulatory, and technical pillars on a shared four-axis template with bidirectional arrows and a layered topology made it possible to identify mismatches between perceived and actual data flow on a case-by-case basis. Because the four axes draws on what controllers already document under GDPR, the participant's mental model and the controller's lawful record could be laid out on the same template, which is part of what made the comparison feel tractable. Whether the instrument would hold up with a larger, more varied group of participants, or with cases drawn from outside a Danish regulatory context, is an open question.

What emerged most consistently at the user pillar was not misinformation but an articulated reluctance to engage further. Most participants reasoned about the first step of the chain and stopped there: only three of seven extended the model to a second-layer recipient, and none placed Meta in the Cookies case, even though several had marked *Received from third party* on the primary card. Where the technical chain was opaque, brand trust appeared to substitute for knowledge, and the regulatory framing most directly relevant to Case C, the Article 9 special-category status of chat content, was not reached by any participant. One participant described the stance themselves as *lykkelig uvidenhed*, blissful ignorance, offered as an active strategy rather than a confession. This is consistent with what Westin (Westin, 1967) would describe as a weakening of reserve, and with the structural opacity of the onward chain that Barth, Ionita and Hartel (Barth et al., 2022) identify as under-covered in existing privacy disclosures. We do not claim more than that the pattern is consistent with these accounts; the sample is too small to make stronger claims.

Two methodological caveats temper the findings. The over-selection visible on the primary card may indicate high privacy concern, but it may equally reflect a confirmation-driven response to a checklist format that made categories easier to mark than to name from memory; we cannot fully separate the two from the data we have. The absence of second-layer cards is perhaps the more interesting signal, since it is what participants did not do when the option was available, but even there we should be careful: the visual language itself may not have been explicit enough to invite participants to extend the chain further, and additional prompting from the moderators in some sessions may have shaped what was placed and what was not.

Within the scope of seven non-technical participants, one mock site, and three Danish DDPA anchors, the framework appears to offer a workable starting point for surfacing the gap between perceived, regulatory, and technical privacy in a single shared vocabulary. The findings here are best read as exploratory rather than conclusive, and the most useful contribution may simply be that the gap seems shaped less by what users get wrong about the first step than by what they have, for understandable reasons, given up trying to see beyond it. Whether this holds more broadly is something a larger study would need to examine.

References

Bandara, R., Fernando, M., and Akter, S. (2019). Explicating the privacy paradox: A qualitative inquiry of online shopping consumers. *Journal of Retailing and Consumer Services*, 51:101947.

Barth, S., Ionita, D., de Jong, M. D. T., Hartel, P. H., and Junger, M. (2021). Privacy rating: A user-centered approach for visualizing data handling practices of online services. *IEEE Transactions on Professional Communication*, 65(1):1–20.

Barth, S., Ionita, D., and Hartel, P. (2022). Understanding online privacy: A systematic review of privacy visualizations and privacy by design guidelines. *ACM Computing Surveys*, 55(3):63.

Bennett, C. J. (2011). In defence of privacy: The concept and the regime. *Surveillance & Society*, 8(4):485–496.

Culnan, M. J. and Armstrong, P. K. (1999). Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization Science*, 10(1):104–115.

Datatilsynet (2021a). Tilsyn med Dating.dk. `https://www.datatilsynet.dk/afgoerelser/afgoerelser/2021/sep/tilsyn-med-datingdk`.

Datatilsynet (2021b). Utilstrækkelige sikkerhedsforanstaltninger hos T. Hansen. `https://www.datatilsynet.dk/afgoerelser/afgoerelser/2021/nov/utilstraekkelige-sikkerhedsforanstaltninger-hos-t-hansen`.

Datatilsynet (2022). Alvorlig kritik af Salling Group for at opbevare passwords i klartekst. `https://www.datatilsynet.dk/afgoerelser/afgoerelser/2022/jul/alvorlig-kritik-af-salling-group-for-at-opbevare-passwords-i-klartekst`.

Datatilsynet (2023a). Alvorlig kritik og påbud til Boligportal for brug af Facebook Business Tools. `https://www.datatilsynet.dk/afgoerelser/afgoerelser/2023/apr/alvorlig-kritik-og-paabud-til-boligportal-for-brug-af-facebook-business-tools`.

Datatilsynet (2023b). Texas Andreas Petersen indstilles til bøde for ulovlig cookie-tracking. `https://www.datatilsynet.dk/afgoerelser/afgoerelser/2023/okt/texas-andreas-petersen-indstilles-til-boede`.

European Parliament and Council of the European Union (2016). Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation). `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679`.

Griggio, C. F., Nelson, B., Sramek, Z., and Askarov, A. (2026). User perceptions and attitudes toward untraceability in messaging platforms. *Proceedings on Privacy Enhancing Technologies*, 2026(1):115–133.

Guo, W., Rodolitz, J., and Birrell, E. (2020). Poli-see: An interactive tool for visualizing privacy policies. In *Proceedings of the 19th Workshop on Privacy in the Electronic Society (WPES '20)*, pages 57–71. ACM.

Milne, G. R. and Culnan, M. J. (2004). Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *Journal of Interactive Marketing*, 18(3):15–29.

Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.

Oates, M., Ahmadullah, Y., Marsh, A., Swoopes, C., Zhang, S., Balebako, R., and Cranor, L. F. (2018). Turtles, locks, and bathrooms: Understanding mental models of privacy through illustration. *Proceedings on Privacy Enhancing Technologies*, 2018(4):5–32.

Privacy Icons Association (n.d.). Privacy icons. `https://privacy-icons.ch/en/download/`.

Schaar, P. (2010). Privacy by design. *Identity in the Information Society*, 3(2):267–273.

Westin, A. F. (1967). *Privacy and Freedom*. Atheneum.

Appendices

Appendix 1: Use of AI

For transparency, and in accordance with IT University of Copenhagen policy, this appendix discloses our use of AI tools during the project.

Claude (Anthropic) was used as a brainstorming and outlining aid throughout the project. We prompted it with questions about the project's structure and content and treated its responses as a starting point, which we then rewrote, expanded, and revised.

Klang.ai was used to transcribe our interview recordings. We uploaded the audio files to the service and used the resulting transcripts as the basis for our analysis.

Claude Code (Anthropic, Opus 4.6 and Opus 4.7) was used for coding assistance in producing design mockups and a small reference prototype for the visual language. Specifically, it generated the interactive single-page web prototype `index.html` (with its companion stylesheet `icons.css`) and the supporting HTML mockups for the cards and context window.

Appendix 2: Prototype

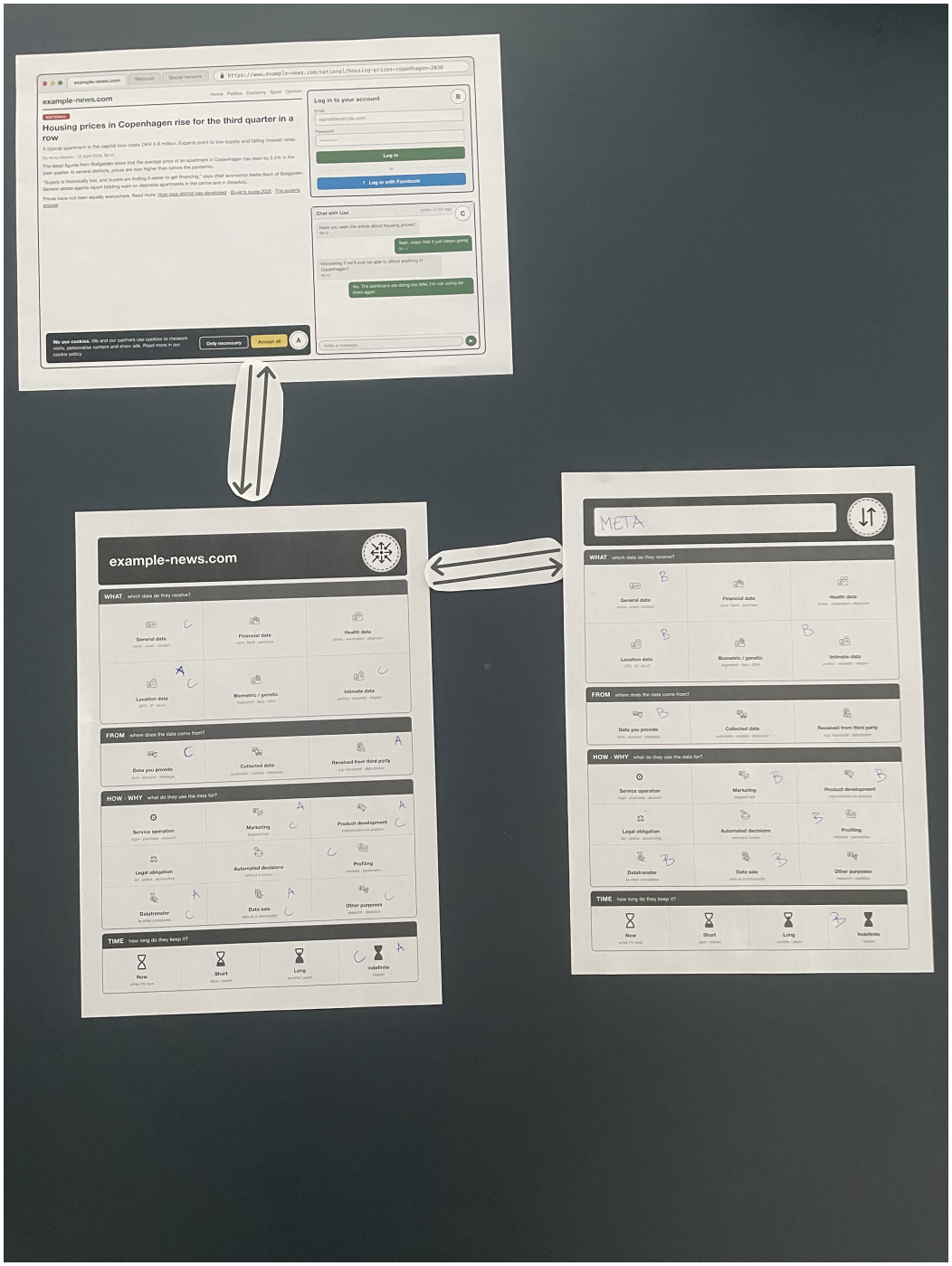
The prototype is available in the project's public GitHub repository at `https://github.itu.dk/adot/Interview-kit-interactive-version`.

Appendix 3: Link to Transcripts

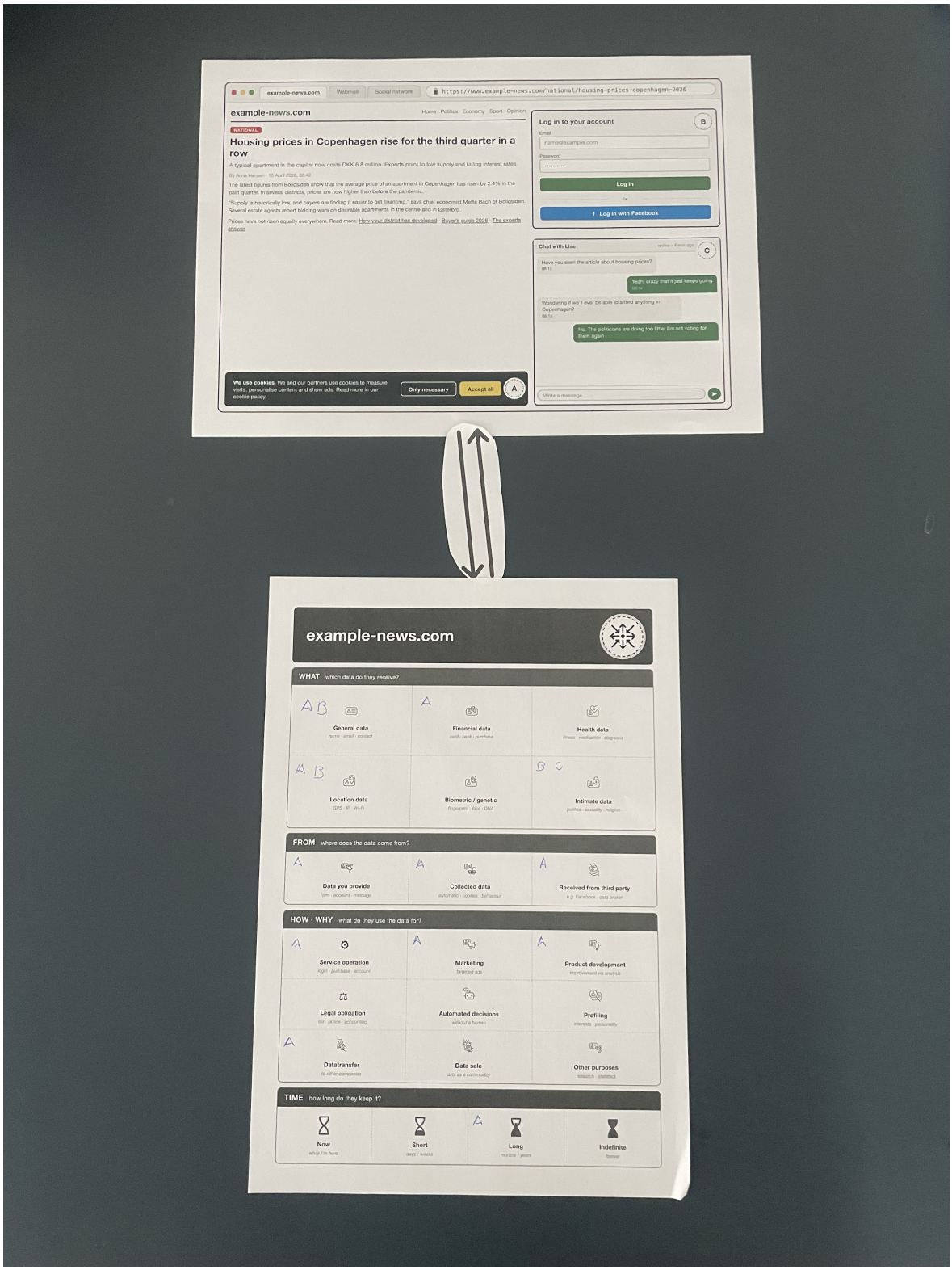
The full AI transcripts are available via the link below.
`https://drive.proton.me/urls/G9KHTX1KT8#hiqW2hSW6fSW`

Appendix 4: Participant Data Flow Cards

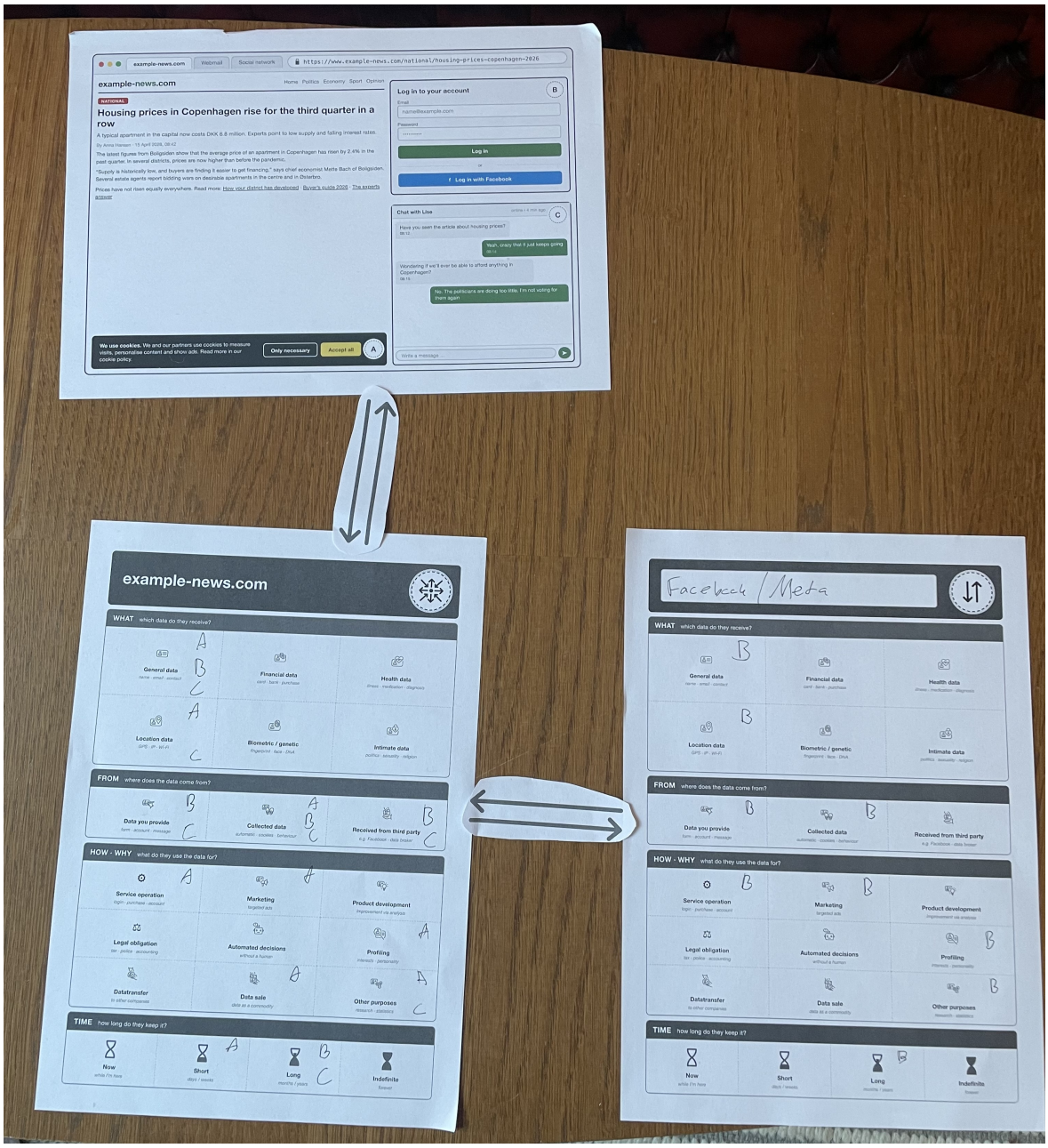
The following pages show photographs of the completed recipient card layouts for each participant. Cards were photographed at the end of each session. Participant codes correspond to P1-P7 as used throughout the analysis.



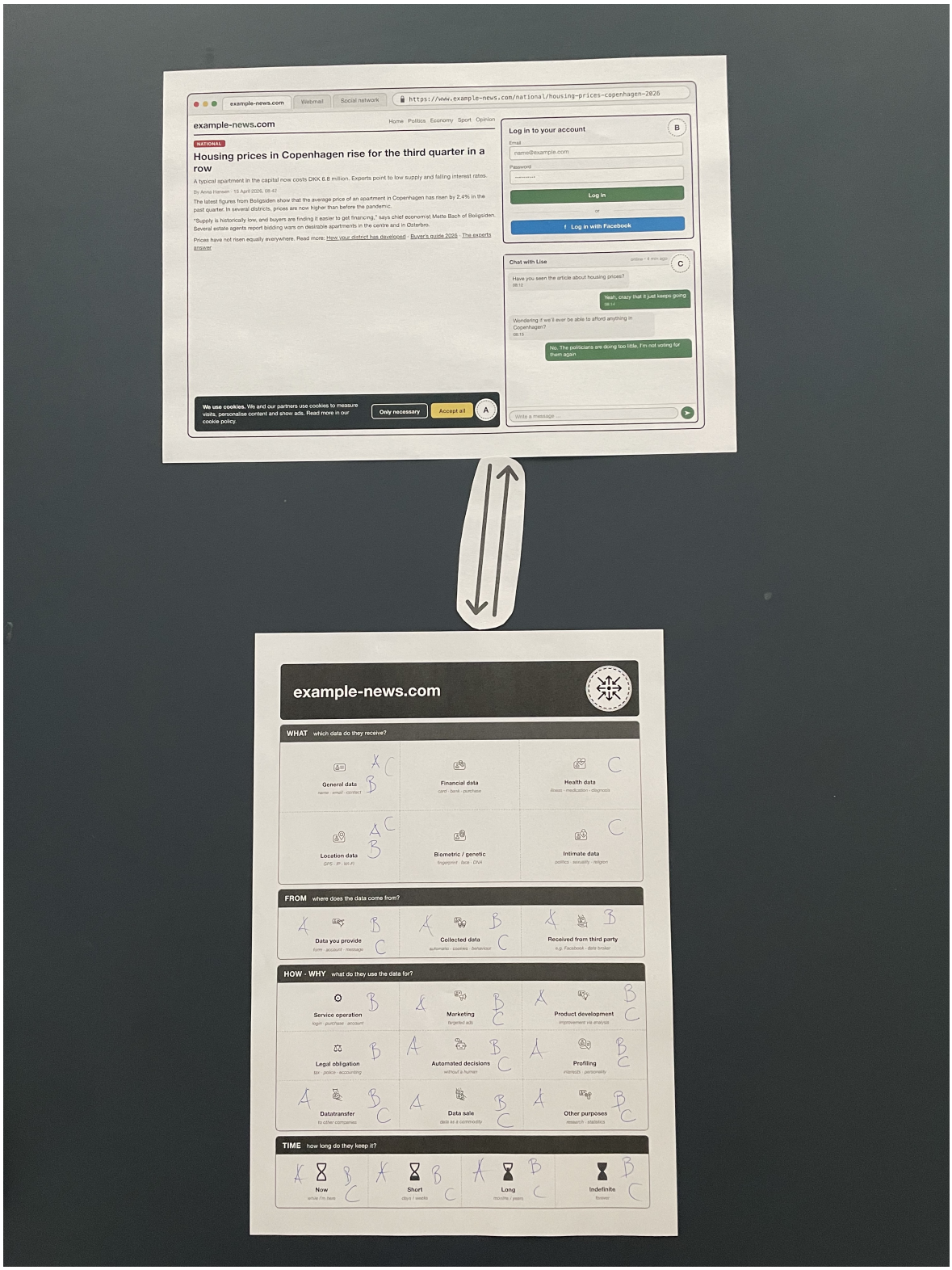
P1: Completed data flow cards.



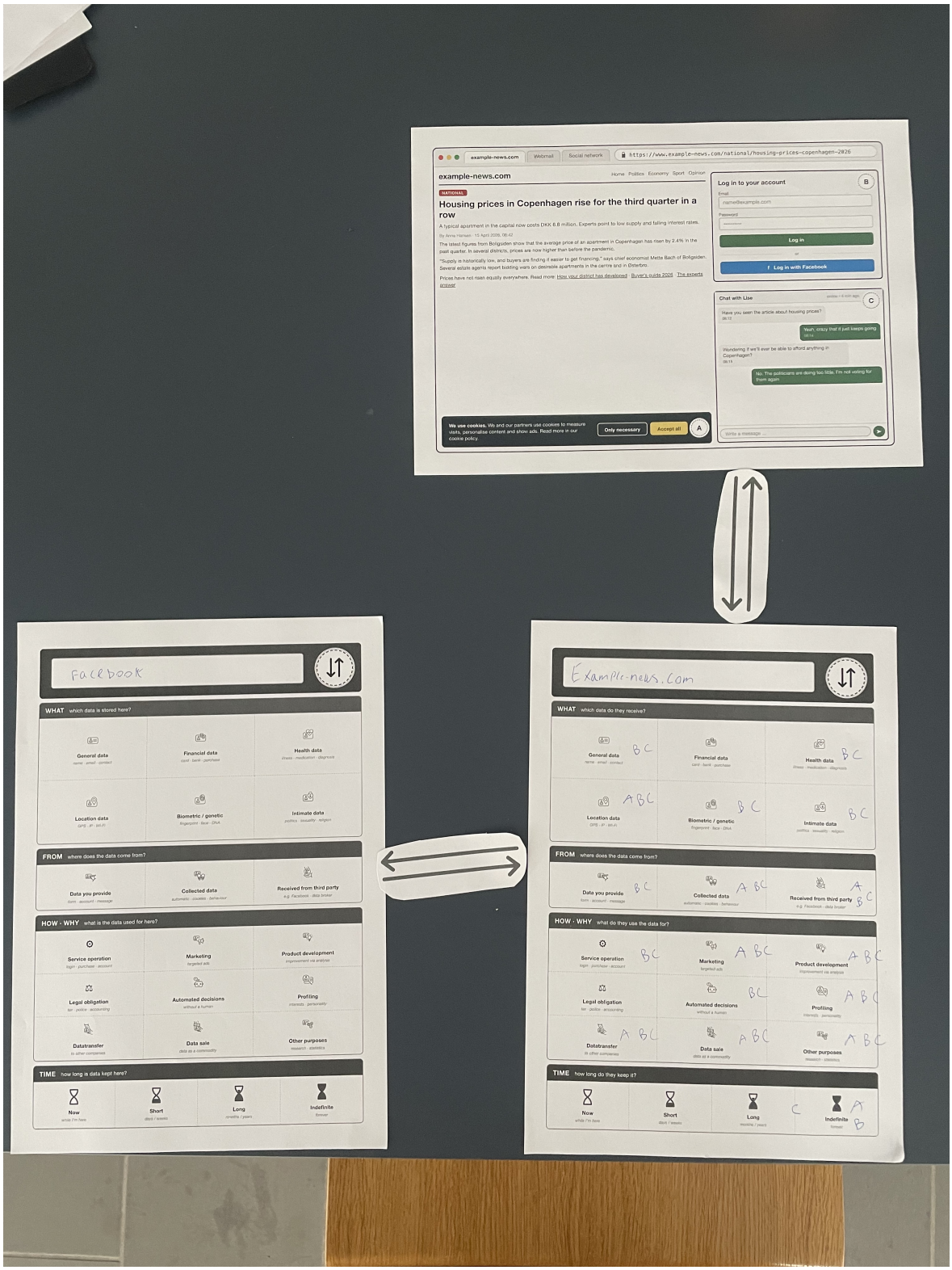
P2: Completed data flow cards.



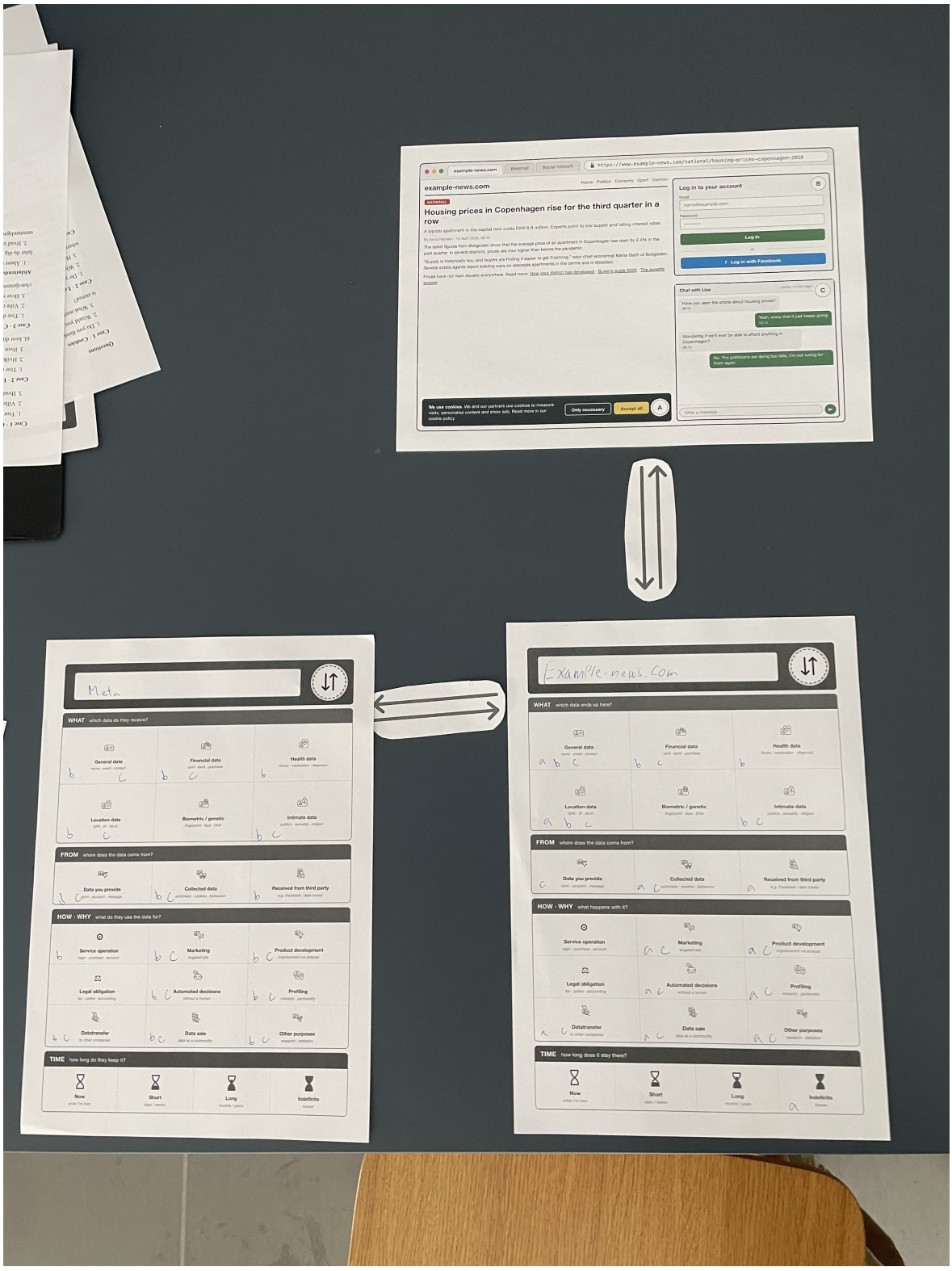
P3: Completed data flow cards.



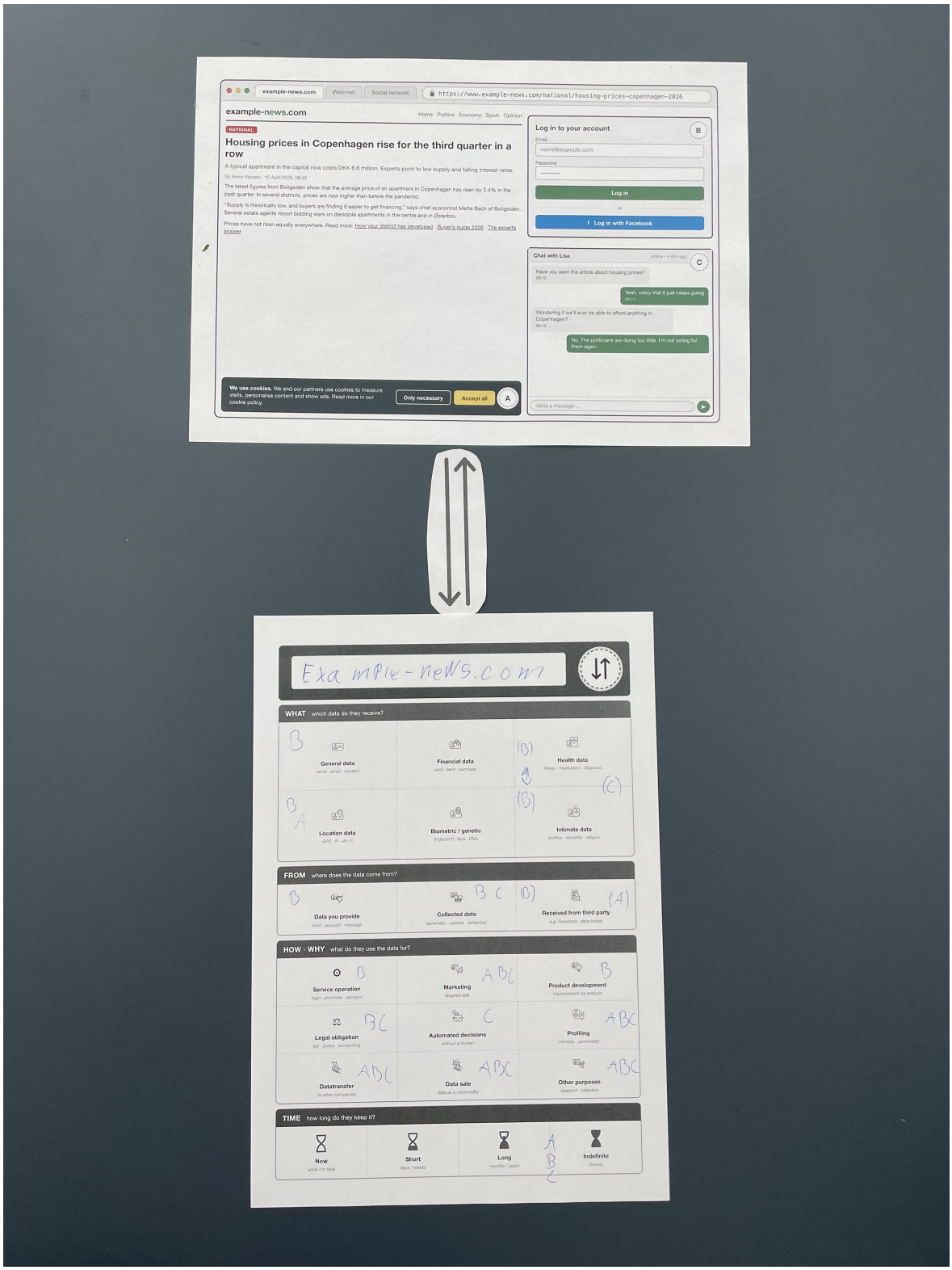
P4: Completed data flow cards.



P5: Completed data flow cards. Note: P5 explicitly asked us to copy all fields marked under Case B on the example-news card across to the Meta card as well, on the grounds that whatever applied to the news site for login also applied to Meta. The Meta card for Case B therefore reflects the participant's instruction rather than independent placement.



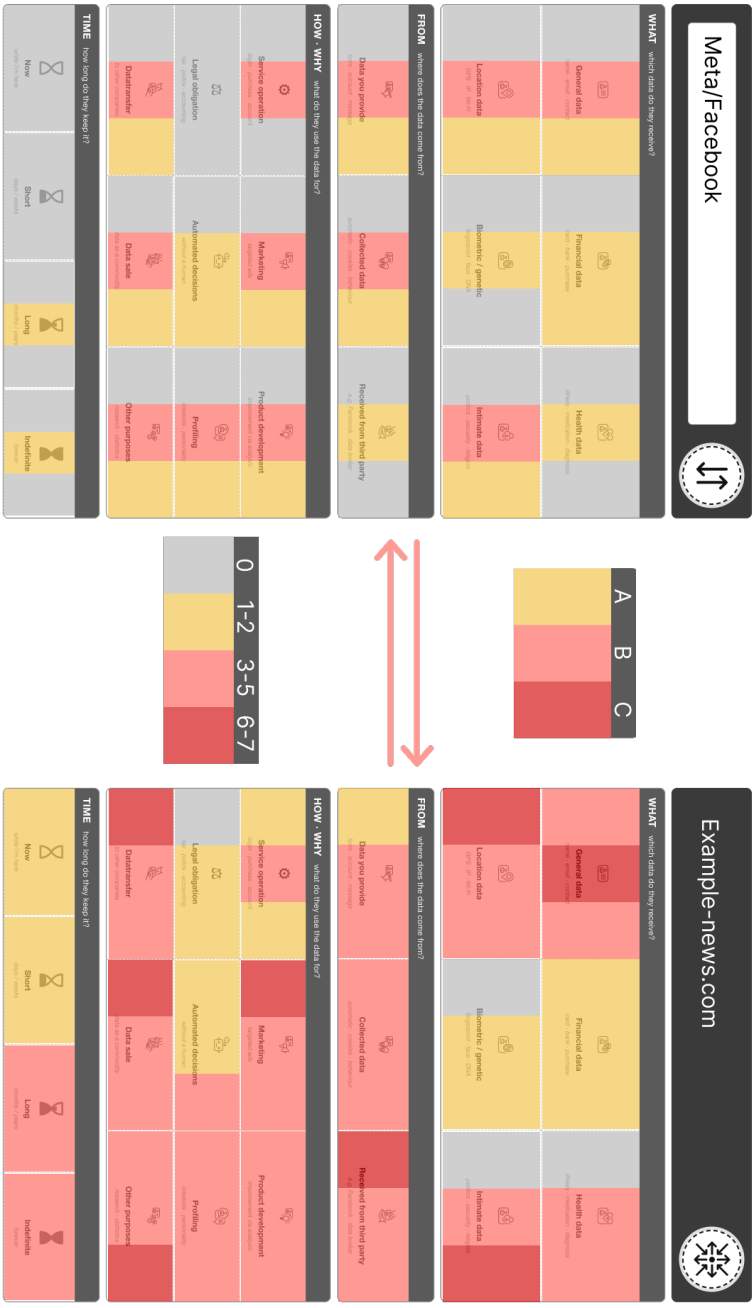
P6: Completed data flow cards.



P7: Completed data flow cards.

Appendix 5: Aggregated Heatmap

The figure below shows the aggregated recipient card responses across all seven participants and all three cases. A full description of how to read the heatmap is provided in the caption of Figure 2 in the Results and Analysis section. The full-resolution version is reproduced here for clarity.



Aggregated recipient card responses across P1-P7 for all three cases.

Appendix 6: Interview Kit

This appendix reproduces the full interview kit used in the seven intercept sessions. It collects, in the order they were used during a session: the recruitment pre-screen handed to candidates before consent, the printed `example-news.com` context window with its three case anchors, the data flow card templates (Core card and blank flow card), the bidirectional arrow set, and the moderator's follow-up questions for each case. Together these are the materials referenced as the interview kit throughout the Methodology and Results.

A. Recruitment pre-screen

Candidates went though this one-page pre-screen shown below with the moderators before any consent or recording.

Q1. Which age group do you belong to?

☐ Under 18 ☐ 18–29 ☐ 30–44 ☐ 45–59 ☐ 60+

Q2. Gender

☐ Woman ☐ Man ☐ Other ☐ Prefer not to say

Q3. How often do you use a laptop or PC, privately or for your studies (so not just your phone)?

☐ Daily ☐ 3 to 5 days per week ☐ 1 to 2 days per week ☐ Less often or never

Q4. Have you ever studied or worked professionally with any of the following?

☐ Computer science / software development / informatics

☐ IT security / cybersecurity

☐ Data protection / GDPR / compliance

☐ Digital marketing / SEO / analytics

☐ No, none of the above

Q5. Are you typically the one friends or family call when their computer or phone is acting up?

☐ Never or rarely

☐ Sometimes

☐ Often

☐ Always, the one who always helps

Q6. Have you ever done any of the following on your own on your computer?

☐ Installed a different operating system than the one the computer came with (e.g. Linux, a full Windows re-install)

☐ Used the terminal or command line regularly for your own tasks (so not just a single copied command from a guide)

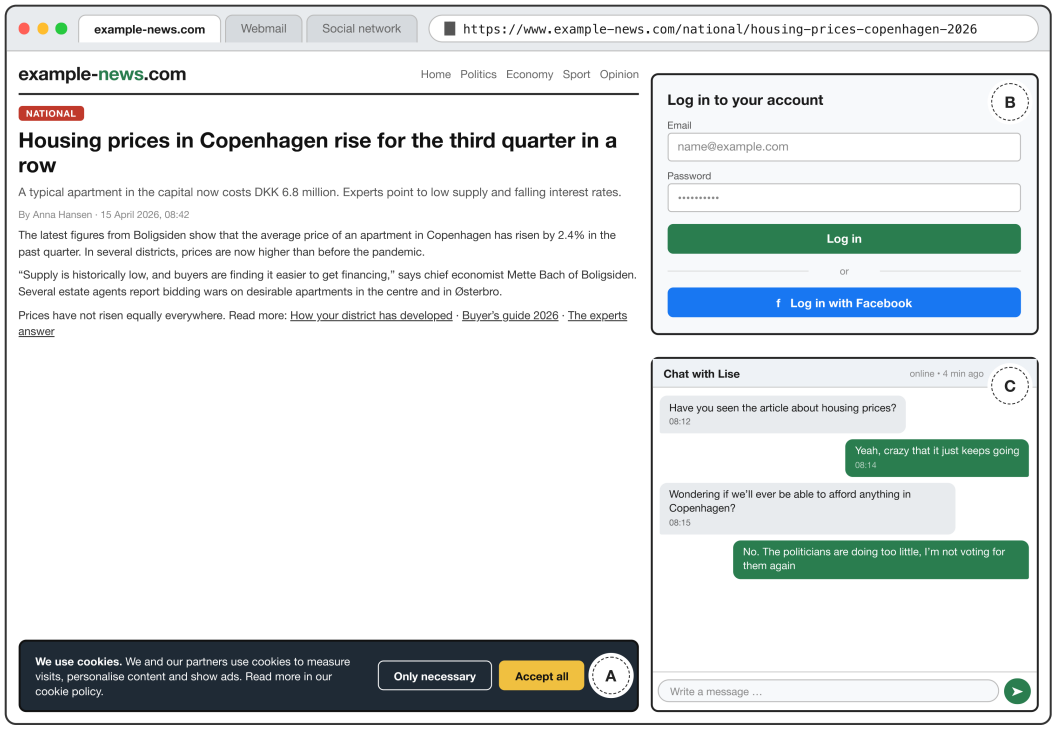
☐ Set up a home server or self-hosted service (e.g. Raspberry Pi with Pi-hole, Nextcloud, Home Assistant, Jellyfin)

☐ Coded a program or a website from scratch yourself (e.g. HTML/CSS, Python, JavaScript). Drag-and-drop tools like Wix, Squarespace, Wordpress themes and Excel formulas do not count.

☐ None of the above

B. Context window (`example-news.com`)

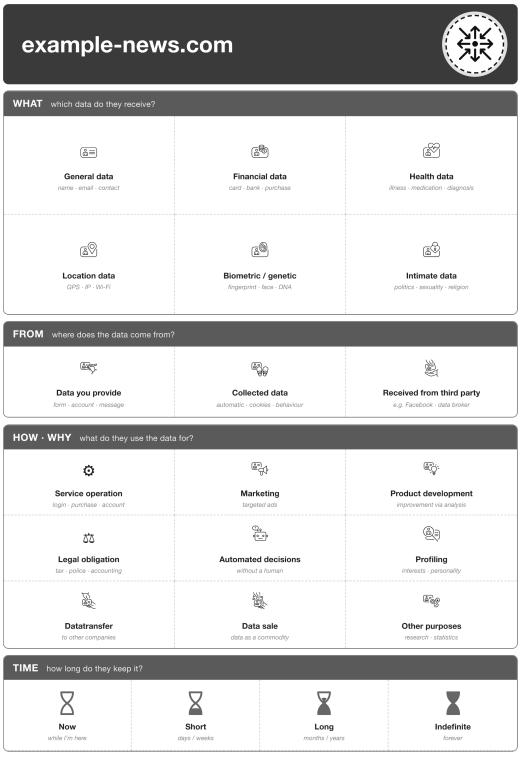
The context window was printed as a single sheet and placed in front of the participant for the entire session. The three anchor points (**A** cookies, **B** login, **C** chat) correspond to the three cases. The Core card was placed on the table at the start of each session, already connected to the context window by one bidirectional arrow.



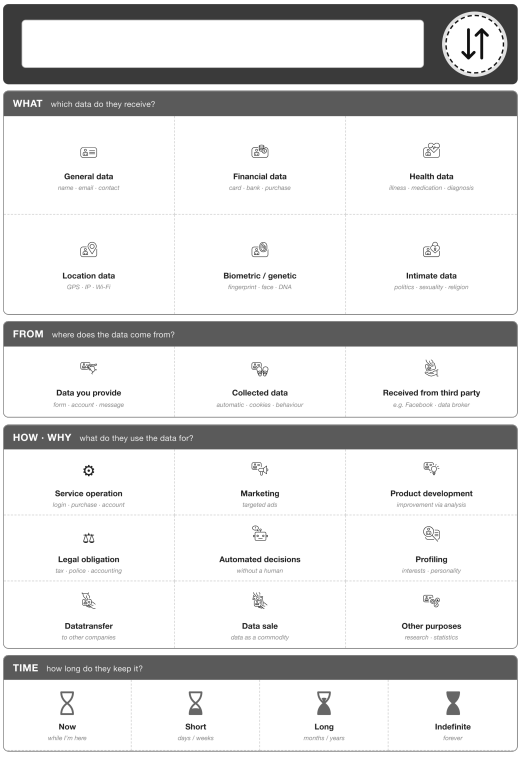
Printed context window for `example-news.com` with the three case anchors A (cookies), B (login), and C (chat).

C. Data flow cards

Two card types were used. The Core card was pre-named `example-news.com` and placed at the start of each session. The other cards were blank data flow card templates that the participant filled in by writing the recipient's name.



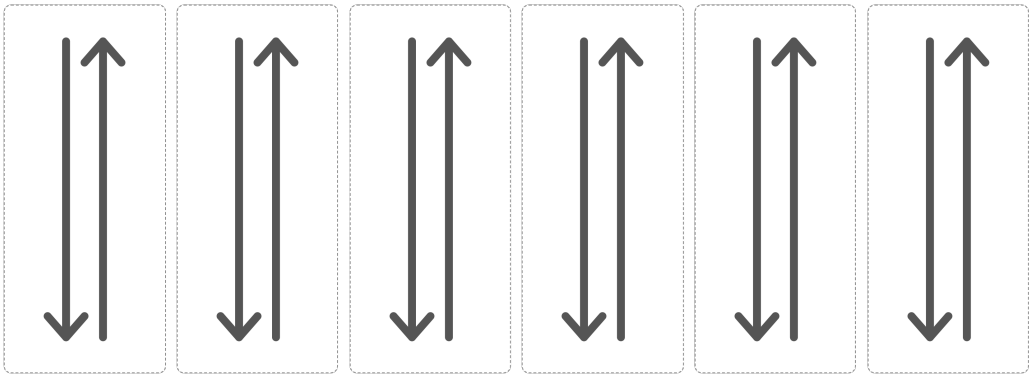
Core card, pre-named `example-news.com`.



Blank flow card template.

D. Arrows

Bidirectional arrows in three lengths were provided so participants could connect cards across the Core, 2nd, and further layers of the onion topology.



Bidirectional arrow set used to connect cards.

E. Moderator's follow-up questions

The semi-structured format meant that each session followed the same structure and covered the same three cases, but the moderators could follow up on interesting responses. The questions below were the prompts the moderators kept on hand for each case, used only where the participant had not already covered the ground.

Case 1 · Cookies

1. Tror du, det gør nogen forskel, hvad du vælger på cookie-banneret?
2. Ville du bruge tjenesten, hvis du vidste præcis, hvad der sker, når du klikker "Acceptér alle"?
3. Hvad vejer tungest for dig her: at kunne læse artiklen uden bøvl eller hvor dine data ligger?

Case 2 · Login

1. Tror du, det gør nogen forskel, hvilken login-metode du vælger?
2. Hvilken login-metode ville du foretrække?
3. Hvor meget betyder bekvemmelighed (hurtigt login, ingen nye passwords at huske) i forhold til, hvor dine data ligger?

Case 3 · Chat

1. Tror du, det gør nogen forskel, hvilken chat-tjeneste du skriver i?
2. Ville du bruge tjenesten, hvis du vidste præcis, hvem der kunne læse beskederne?
3. Hvor meget betyder bekvemmelighed (at skrive direkte på siden frem for at skifte til en anden chat-tjeneste) i forhold til, hvor dine data ligger?

Closing

1. Åbent for al feedback: Hvad overraskede dig, hvad fik dig til at tænke anderledes, og hvordan føler du dig nu sammenlignet med, inden vi startede?
2. Hvad tænker du om denne visuelle måde at forstå, hvor ens data ender - for eksempel sammenlignet med at læse en privatlivspolitik?