Understanding and Designing Human Data Relations

Alex Bowyer

5 Case Study Two: The Human Experience of GDPR

[intro 2 paras highlight what I am doing in this chapter]

[condense from intro text] The world is data-centric; our

everyday lives have a digital side, and organisations that provide services to us collect our personal data, often without fully informed consent, in order to support algorithmically-driven business decisions. Data has become a commodity, exploited and traded for commercial advantage and for behavioural insights that often serve advertisers more than users. Personal data thereby becomes an asset of providers, which the users that the data concerns cannot easily see or access. Nor can users see and understand how their data is being used. These issues fundamentally affect the experience of using digital services, and therefore good Human-Data Interaction [79] should be a core matter of concern for HCI and UX professionals [80]. There is power imbalance over personal data [48–51]; it is currently scattered and trapped beyond our individual reach. In 2018, the European Union's General Data Protection Regulation (GDPR) [36] came into force, as an attempt to rebalance power by granting individuals rights to access their data and have its usage explained. In the three years since, similar policies have been introduced around the world, and there has been a visible impact upon providers, who may be required to respond to requests for data access and provide people with a copy of their data. While the GDPR has become a valuable tool for transparency that has been examined and harnessed by various researchers, few have done so from a user-centric, human-data interaction (HDI) [80] perspective. Our research seeks to go deeper than prior work [4] by examining the human experience of using one's GDPR rights: exploring compliance, response quality, individual attitudes to data-holding organisations, and the impact upon the service relationship, in order to identify challenges, inform individuals' choices around data and explore how data policies and practices could be redesigned. This paper presents qualitative and quantitative findings from several months of in-

depth analysis of interview transcripts, recordings, sketches and spreadsheets, delivering insights for policymakers, data-holding companies and individuals. We find that the GDPR's aim to provide individuals with control over and value from their data is being hampered by non-compliance and poor-quality data responses, and that both data holders and individuals stand to gain from improved data access and transparency. Informed by our findings, we recommend specific design approaches that can inform the different parties involved. Our key contributions include:

- A detailed account of the extent to which service providers complied with GDPR requests, including types of data that were received, as well as perceived completeness, accuracy and usability
- A summary of the hopes, plans and imagined uses of personal data that motivate GDPR requests
- Thematic findings showing that data holders were seen to be evasive and individuals struggle to make sense or use of their returned data
- Evidence of a detrimental impact upon individuals' evaluations of the integrity and trustworthiness of data-holding organisations as a result of scrutiny of privacy policies and GDPR responses
- Sociotechnical design insights to redesign the GDPR for greater effectiveness and how companies could rethink the ways in which they structure their users' relationships with personal data
- Advice for individuals on how to use the GDPR to positive effect.

5.1 Context: Accessing Your Personal Data Using Your GDPR Rights

2 BACKGROUND To ground our exploration of GDPR we begin by presenting some general context, then we address research that has explored GDPR and its limitations, before positioning this work in relation to more human-centred studies of Human-Data Interaction and personal data ecosystems.

5.1.1 GDPR: Legislation Seeking to Empower Individuals in a Data-Centric World

The widescale adoption of personal computers [108] and smartphones [22,41] combined with the advent of cloud-computing [65] have led to ubiquitous storage of personal data [49,57]—data about people—by service providers across all walks of life in both commercial and public sectors. Data-centric companies now dominate almost every sector [23,68]; we live digital lives [14,123] and the collection of our personal data has become inevitable. This data is increasingly used to make decisions, being seen as a resource which can be mined for value and exploited at scale [19,35,84,96,106]. Unfortunately, people have little awareness of—or access to—this data; the large-scale data-centric systems that drive modern life largely function as opaque 'data traps' [1] and data collection is often unwitting [95]. The World Economic Forum's "Rethinking Personal Data" project recognised the critical role that data now holds, and identified that "an asymmetry of power exists […] created by an imbalance in the amount of information about individuals held by industry and governments, and the lack of knowledge and ability of the same individuals to control the use of that information" [48–51]. Data becomes a proxy for people's direct involvement [18] and without effective data access [44] people are disempowered, lacking agency and control of data that is held about them [31,79]. Since the 1970s, laws aimed to protect individuals'

data and data rights [15,29,86,109,121], but up to now such efforts have been ‘almost useless in limiting the growth of surveillance’ [77]. The EU’s GDPR [36] with its designed-to-hurt fines that are now being implemented [21,27,63,67] for non-compliance has finally begun to have some impact [7], giving at least 513 million people rights to timely data access, explanation, erasure and correction [54]. It is regarded as a landmark piece of legislation and a strong template for individual data protection, having inspired similar legislation such as the CCPA in California [64] and others in India, Japan, Turkey and beyond. A focus on data protection has become particularly important and gained increased global public awareness due to the Snowden revelations [39], Cambridge Analytica scandal [122] and the COVID-19 pandemic [45,85]. Since the GDPR’s launch in May 2018, many large consumer companies have developed ‘privacy hubs’ or improved privacy policies where users can learn how their personal data is handled or easily download copies of it [110–113]. However, we do not know how effective these measures are for users; i.e. whether—by compelling data holders to create such offerings and respond to access requests—GDPR succeeds in its goals to ‘enhance the data protection rights of individuals’ [30] and to give people ‘control over their personal data’ [94] to redress the aforementioned power imbalance between data holders and individuals .

5.1.2 Current GDPR Research and its Limitations

Since 2018, the adoption of the GDPR has opened up new possibilities for research [28]; the ability to obtain one’s data records from organisations provides the general public with a potential deeper view inside those organisations. This legally-enforced transparency can also provide researchers with a

window into organisations and their processes that was previously only available based on goodwill. Ausloos and Veale [8,10] provide an outline approach for such research and discuss ethical and methodological considerations. The GDPR process itself has also been examined from many perspectives: to understand data holder's compliance with legislation [6,9]; to evaluate data portability [102] and 'privacy by design' [99]; to compare its effectiveness in public/private sector contexts [88] or in improving explainability [45], fairness [60], consent [52], transparency [91] and the reduction of data breach risks [43]. Potential negative impacts have also been considered; the threat to privacy [20] and potential impediments to health research [26]. The impact of the GDPR spans legal, social and technology domains. Yet, there is scant research into the individual human experience of the GDPR. Alizadeh et al. conducted a study with 13 users of a German loyalty programme and interviewed them before, during and after they made GDPR data requests [4], finding better responses and GDPR education were needed. This study was limited in breadth (targeting only one provider) and depth (returned data was discussed largely at a high level of 'were your expectations met?' and did not examine potential data use benefits). The impact upon the participants' relationship with the provider was not explored. Recent work [20,42,107] has established that openness and transparency around data handling are key to services establishing users' trust; in a commercial context this impacts customer satisfaction and business success. There is a need to understand the experience people have when using the GDPR; companies' GDPR processes have been designed to comply with litigation rather than focusing on user needs or desires [3,74,104]. Understandings on 'the human experience of the GDPR' could inform the design of improvements to digital GDPR

mechanisms and help identify potential policy improvements.

5.1.3 Human-Data Interaction: Towards a Human-centric Personal Data Ecosystem

In 2017, the average American Internet user had 150 online user accounts with different providers [24]. Data for the UK show the number of service and supply relationships each individual has to manage increasing from around 45 in 1997 to around 250 in 2020 [47]. As the amount of personal data relating to each of us has increased, the need for individuals to be able to manage it has grown. In the 1990s this was considered through the personal information management (PIM) lens of giving people ‘a place for their personal data’ [59] and then facilitating easy filing-and-retrieval to improve task efficiency and personal productivity [5]. Since, there has been growing recognition that this problem needs to be tackled at a “whole life” level–our data exists across devices and across providers [2,83,100]. Through the fields of personal informatics [69,70] and the quantified self [62], researchers and hobbyists have explored the practicalities of collecting and integrating data about oneself so that the user is able to reflect upon it, gain insights and take informed action [70,71,89]. In this wider sociotechnical context, where data is held by multiple service providers offering limited use and access, this challenge increases substantially; data cannot be moved freely [16] and our ability to connect our data is limited [31]. Managing personal information has shifted from ‘arranging data bookshelves’ to becoming a multi-party negotiation over representation, ownership, access and consent [95]. Data is a shared resource with multiple users; few researchers have begun to look at people’s interactions with data in this context [46,87,105] – and not yet in the context of accessing data held

by service providers. Researchers have identified that a human-centric approach to personal data is needed, placing individuals at the centre, as controllers and overseers of their own personal data ecosystem [83,93]. This is an emergent space of much activity and research [13,32,81,82,93,114,124] and provides a strong framing for us to evaluate the human experience of–and interaction with–the GDPR, indicating relevant research questions, such as: From the starting point of users having diminished agency and control over their data [31,103], do the GDPR's access rights provide the effective access [44] people need? Does the GDPR help people to achieve legibility, agency and negotiability, which are needed for effective Human-Data Interaction (HDI) [80]? Does it allow people to have a relationship with rather than through their data? Our study aims to explore these research gaps by documenting the experience of using one's GDPR rights and assessing how well the process meets users' needs and expectations to uncover problems and possible solutions that could address them.

5.2 Study Design and Configuration

To address our research challenges, we conducted 31 qualitative interviews with a convenience sample of 11 individuals from a population of researchers and students at (or connected with) Newcastle University, aged 20-40 years; self-identifying as 5 females and 6 males. Participants were not data experts (only 1 had previously made a GDPR request), but were computer-literate, educated to degree level, and used to reflecting critically on their own behaviours and opinions. Participants were compensated for their time with Amazon vouchers worth £20.

Figure 1: A Journey Map of Each Participant's Study Progression

Each participation progressed at its own pace (see Figure 1) with participants invited to three separate 1-on-1 interviews between December 2019 and April 2020. The scope and purpose of each interview was as follows:

1. Interview 1: Sensitisation, Life Exploration and Company Selection [1 hour, in person]. Participants were sensitised to the research context using an interactive tour of a poster display on the topics of GDPR rights, potential data-holding organisations, potential types of data and potential uses for GDPR-obtained data. Baseline data was collected on participants' hopes and motivations, their current understanding of personal data, data access, data control, and power as it relates to data. Using a sketch interviewing [53] technique, participants mapped out their 'data lives' (e.g. Figure 2), annotating key organisations that they have relationships with, types of data those companies might hold, and feelings about such data use and storage by each holder. Each participant selected 3-5 candidate companies to explore with GDPR requests.

2. Interview 2: Privacy Policy Reviewing, Goal Setting and GDPR Request Initiation [1 hour, in person]. To stimulate reflective thinking and measure impacts, participants were asked to discuss and score their initial feelings of trust and power with each company. Participants then viewed key sections of privacy policies on a screen with the researcher, to identify each company's statements on collection and use of personal data. Participants then initiated an email GDPR request for each company, which had been prepared using a tried-and-tested template

generated by personaldata.io [101]. Interview 2 took place in person, except for P10 & P11 whose interviews took place over Zoom due to the COVID-19 pandemic.

3. Interview 3: Detailed GDPR Response Review [2 hours, online video call] Having allowed sufficient time for GDPR requests to conclude (there is a legal duty to reply within 30 days), a deep dive into the specifics of each GDPR experience took place. Participants' personal data was not collected by the research team; screen sharing was used to show excerpts to the researcher where the participant wished to do so. Participants were asked a structured set of questions about the completeness and value of any data returned, as well as new evaluations of trust and power, whether their hopes had been met, and any general feelings about the experience. Answers were recorded in a screen-shared spreadsheet, which was also used to structure the discussion (for a sample cf. attachments).

Figure 2: An Example Life Sketch from Interview 1, with Data Handling Companies in Red, Data Types in Blue, and Feelings in Green

Interviews were audio and video recorded, then auto-transcribed using Google Recorder/Zoom, producing a 370,000-word corpus. Transcripts were split up by topic and analysed through reductive coding cycles to produce thematic findings (see section 5). Quantitative data from interview spreadsheets was summarised and analysed (see section 4). Sketches, recordings, screenshots and field notes aided interpretation of the transcripts.

5.3 GDPR Request Outcomes

4.1 Interview 1: GDPR Target Selection Initially eight participants chose 5 companies and three chose 4 to request data from. One participant (p9) withdrew from the study due to COVID-19 after Interview 1. Five participants withdrew a chosen company upon further consideration. Reasons for withdrawing chosen targets included having one's personal data mixed with other household members (Netflix, Morrisons), not wishing to impact active customer support matters (LNER), and inability to contact the provider by email (see below). One participant selected Newcastle University, which was vetoed by the research team to avoid conflicts of interest. Hence, 41 out of a possible 52 GDPR subject access requests were made (to 28 distinct data holders; cf. Table 1): Table 1: Types of Data Holding Organisation Targeted for GDPR Requests by Study Participants Type of Company Company Namesa Major Internet Companies Apple (3), Amazon, Facebook (4), Google (5) Hardware Companies Apple (3), Huawei, Google(5), Philips Hue (smart lightbulb manufacturer) Online Platforms/Websites Airbnb, Bumble (dating site), Check My File, Credit Karma, Direct Line, last.fm, LinkedIn Social Networks & Dating Facebook (4), Instagram, LinkedIn, Bumble (dating site) Software/App Manufacturers Freeprints, Niantic (creators of Pokémon Go), Natural Cycles (a menstrual tracker), Revolut, Spotify Transport Companies Tyne Tunnels, Nexus (Tyne & Wear Metro), LNER Retailers & Loyalty Schemes Amazon, Tesco, Sainsbury's, Nectar Telcos Virgin Media, Three Sports Clubs Sunderland AFC a Where a company was chosen by more than one participant, the number of participants choosing that company is shown in brackets. To ensure fairness and consistency, the aim was that all GDPR requests be sent by e-mail to the identified Data

Protection Officer, requesting both a subject access request [55] and a data portability request [56] be initiated, and specifically enumerating and asking for those datapoints that the company stated in its privacy policy, as well as those which the GDPR entitles individuals to obtain. To identify these datapoints, company privacy policies were analysed and the necessary information was compiled in personaldata.io's semantic wiki [115] which has a feature to generate bespoke GDPR request emails, which we adapted and used (cf. attachments). Facebook, Apple, Huawei and Philips Hue do not offer a contact e-mail address, so the email text (shortened where length restrictions applied) was pasted into a contact form. In one case, entertainment website ifun.tv, the only available means of contact was via WeChat, resulting in the participant (a Chinese citizen) choosing not to contact ifun.tv due to fear of Chinese government surveillance. From our analysis of these companies' and others' privacy policies and with reference to GDPR rights, we constructed a taxonomy of the types of personal data that could be returned, using terms from those privacy policies and GDPR legislation: there are five types of personal data, as shown in Table 2.

Table 2: Types of Personal Data Potentially Accessible from Data Holders via GDPR Rights Type of Personal Data Description Examples Volunteered Data Data that the data subject has directly provided to the company through upload, contact or form completion. Personally Identifiable Information (PII), contact details, user-generated content, photos, files, profiles, settings, communication history, financial information, security credentials, surveys/forms. Observed Data Data that has been indirectly or automatically collected about the data subject through product/service use or customer/staff interaction. App usage information, behaviour on website,

search/browse history, location tracking/tags, activity/health tracking, technical/device information, network/telco/ISP information, cookies & pixel trackers, staff observations, customer interaction notes. Derived Data Inferred data or profiles that have been created through algorithmic or human analysis of volunteered, observed or acquired data. Interest profiles, advertising demographics, market segmentation, customer categorization, product/service recommendations, internal customer codes. Acquired Data Data that has been obtained or purchased from external sources such as civic records, reference agencies, advertisers or third parties. Public records and information from internet searchs, reports or reviews from individuals, electoral roll data, credit checks, fraud checks, criminal record checks, e-mail/interest lists from advertisers, information shared between affiliates, sister companies or partner organisations. Metadata Information about how the other four categories of data have been handled, including storage, processing, uses, decision-making and external sharing. Names of third parties data has been shared with, details of where data is stored and when/where it has exited the EU, explanations of how data has been used in automated or human decision making, legal bases for storage and processing. 4.2 Interview 2: Privacy Policy Review and Goal Setting Participants reviewed and discussed privacy policies for their chosen target companies and were asked to define hopes and expectations for each GDPR request (see Table 4). These most commonly related to seeing the breadth and depth of data collection by companies, understanding what was being inferred and how personal data was used, and to use such information to better assess trustworthiness of those companies. Other motivators included the desire to reflect on one's own past data to gain self-insight, and to take control of or delete held data. Minor motivators included

learning, creativity, fun, nostalgia, curiosity and the desire to shed light on specific incidents or answer specific questions.

Figure 3: A Sankey diagram giving an overview of the GDPR requests (R) undertaken by our participants (P). At the conclusion of interview 2, participants were provided with the emails and instructions to start their GDPR requests, which progressed as illustrated in Figure 3. Eight requests resulted in no data being obtained, due to either data holder non-responsiveness, inability to access the right account or satisfy ID requirements, or confirmation being received that there was no data to supply. 32 requests (80%) resulted in at least some data being returned; 10 of these directed the participant to use a publically-available download dashboard such as Google Takeout, and the rest resulted in data being made individually available. Of these, one was posted as printouts, another was posted on CD-R, and the rest were delivered by email (sometimes involving a secured online website to download). While 22 companies supplied bespoke data packages, 4 did not return it within the 30 days the legislation specifies (note: requests took place within the context of a global pandemic so response rates may not be typical). Following discussion, all 32 requests receiving data were considered not to have returned all requested data across all five of the categories in Table 2.

4.3 Interview 3: Reviewing the GDPR Response Once each participant's GPDR requests had reached a conclusion point (as described above), they were invited to discuss the GDPR response in detail. Participants were asked to describe (and optionally show) the data they had received, then to evaluate the data holder's response for each data type, according to multiple metrics designed to assess the perceived quality of the GDPR request handling and the subjective value of any returned data. All questions were posed from the perspective

of (a) the data that providers said they collect and process in their privacy policies, and (b) the rights that the GDPR specifies, to ensure discovery of missing data or unfulfilled rights would be considered objectively. Participant responses were considered quantitatively (cf. Table 3) and qualitatively (cf. section 5). Table 3: Presence and quality assessments of GDPR responses by data type (as percentagesa) Type Valued? b Returned? Complete? Accurate? Understandable? Meaningful? Usable? Useful? Derived 82% 39% 10% (dk:13%) c 20% (dk:20%) c 40% (p:40%) d 40% 0% 20% Acquired 81 49 16 (dk:16) 50 (dk:25) 75 (p:0) 50 25 17 Metadata 73 4 0 (dk:7) 0 (dk:0) 0 (p:100) 0 0 0 Volunteered 57 53 55 (dk:0) 92 (dk:0) 72 (p:20) 72 52 58 Observed 48 33 18 (dk:12) 57 (dk:30) 61 (p:20) 57 52 61 a Percentages represent the proportion of “Yes” answers to each question, per data subtype, from all those where a judgement was given. b Participants were asked whether this category of data from each provider would be valuable if they were to receive it. b dk = don’t know (percentage of cases where participants felt unable to assess data accuracy or completeness). c p = partially (percentage of cases where data was judged partially understandable). Table 3 shows quality assessments for each data type, with rows descending by subjective value. Notably, the kinds of data participants value most (derived, acquired and metadata) were less frequently returned, especially metadata (returned in 4% of cases). Where data was returned in these categories, it suffered from poor data quality, often judged as incomplete, inaccurate, unusable and not useful (although acquired data was largely understandable). At 53%, even the most returned category, volunteered data, was lacking. Where it was returned, accuracy (92%), meaningfulness (72%) and understandability (72%) were high.

Observed data was least valued and also rarely returned or complete (yet judged to be of moderate quality). Across all data types, we note that data was only judged to be complete in 22% of cases, and in 62% of cases personal data specified in privacy policies to be collected was not returned, despite the legal obligation. When invited to revisit their hopes and anticipated data uses to conclude the third interview (Table 4), participants felt goals were not fully met in 78% of cases, and 54% were not met at all. Specific problem areas include (1) the desire to understand what providers infer from held data (7 participants), which was unmet in 73% of cases and only fully met in 7% of cases; and (2) the desire to delete one's data, which occurred in 10 cases but was only met in one of them. Four wholly unmet hopes were to investigate specific incidents (GDPR responses were often delivered as a one-off package without any backchannel), to secure data, to check accuracy, and to move data to another service. Table 4: Participants' hopes, imagined data uses and goals for GDPR, as well as resultant outcomes Hope or Goal Distinct Distinct Specific companies Was this hope met (%)? instances of this goal participants in mind for this goal, if any Unmet? Partially met? Fully met? Understand the breadth and depth of what data is collected 24 7 Amazon, Apple, CheckMyFile, Credit Karma, Facebook, Google, LNER, Nectar, Philips Hue, Spotify, Tesco, Three, Virgin Media 42% 17% 42% Understand what is inferred about you from your data 15 7 Amazon, Apple, Direct Line, Google, Instagram, last.fm, LNER, Spotify, Tesco, Three 73 20 7 Reflect on past activities & gain insights 14 5 AirBNB, Apple, Google, last.fm, LNER, Tesco, Virgin Media 57 36 7 Assess provider trustworthiness 12 6 Apple, Credit Karma, Direct Line, Facebook, Freeprints, Nectar, Niantic, Sunderland AFC, Tesco, Three 42 42 17 Remove your data & control/limit its use 10 3 Bumble, ifun.tv, Instagram 90 0 10 See inside

‘black box’ algorithms & processes 9 4 Amazon, Facebook, Google, Tesco 56 11 33 Find patterns/habits & track goals 6 5 last.fm, Nectar, Spotify, Tesco 17 50 33 Understand how and why your data is used 6 5 Direct Line, Google 50 33 17 Investigate specific questions or incidents 4 4 AirBNB, Three, Credit Karma, Instagram 100 0 0 Play with, create, hack & remix your data 3 3 Google 67 0 33 Nostalgia, fun & inspiration 3 3 Spotify, Niantic 33 33 33 OVERALL 18 goal types 10 people – 54% 24% 22% For unabridged table, including goals held by fewer than 3 individuals, cf. data attachments. 4.4 Perceived Power and Trust We examined how participants’ feelings towards the data holders changed throughout the process. Participants were asked to assess trust from 0 (total distrust) to 10 (total trust); and were asked to assess power on a scale of -5 (total provider power) through 0 (balanced power) to +5 (total individual power) and to explain their reasons for initial rating and for any change. By repeating the same question at different times, we were able to observe changes in attitude; these changes are summarized in Figure 4. Many participants’ attitudes did change as a result of the experience, for both perceived power (45% of cases) and trust (66% of cases). Of those with changed attitudes, the majority were negatively affected; in 29% of cases participants perceived a loss of individual power, and in the majority (52%) of cases, participants felt more distrustful of GDPR targeted companies after completing the process. However, it is important to note that in some cases GDPR had a positive impact; in 17% of cases participants felt their perceived power had increased, and in 14% of cases participants felt more trusting of providers after GDPR.

Figure 4: Distribution of Net Changes in Participant’s Perceived Power and Trust Scores over the Study’s Duration

5.4 Thematic Findings

[description like in c4 plus include some of following] Here we present outcomes from a deeper analysis of the participant experiences summarized above, derived through over 200 person-hours of iterative data analysis [76] of the interview transcripts. Our three key thematic findings are:

5.4.1 Themes & Subthemes

[convert this to prose]

1. **Insufficient Transparency** (Theme 1): Organisations appear evasive over data when responding to GDPR, leaving people “in the dark” even after making GDPR requests.
2. **Confusing Data** (Theme 2): When presented with their data, people struggle to understand it and relate it to their lives and are not able to make use of it.
3. **Fragile Relationships** (Theme 3): Companies’ data practices, and in particular their privacy policies and GDPR response handling, can be impactful to customer relationships, carrying a risk of damaging trust but also the potential to improve relations. These themes are detailed in 5.1, 5.2 and 5.3 respectively.

[insert Table for each of the three themes]

Table X. **Theme 1 - Insufficient Transparency**. Subthemes & Participant Quotes.

Subtheme	Description & Quote	Status

Table X. **Theme 2 - Confusing Data**. Subthemes & Participant Quotes.

Subtheme	Description & Quote	Status

Table X. **Theme 3 - Fragile Relationships**. Subthemes & Participant Quotes.

Subtheme	Description & Quote	Status

5.4.2 Theme 1: Many Companies are Evasive and People are “Still in the Dark”

5.4.2.1 Compliance

Responses to GDPR requests were broadly unsatisfactory (cf. section 4). Cases where data was not returned on time, or at all, were clear breaches of the legislation. Participants saw the incompleteness of returned data as a further failure to comply. At the beginning of the study, participants were reminded of their GDPR legal rights (most were already aware to some degree, as expected for such participants [90]), and several participants referred to these rights when reflecting on their experiences. e.g., P5 used legal rights to gauge the quality of one of her responses: “I feel more concerned now, […] what they’ve given me seemed reasonable. But then comparing against what we asked them for, what I’m legally [entitled to], it’s a fraction.” –P5 The extent to which participants expected their rights to be honoured varied, with some participants expressing scepticism from the outset. To them, poor responses were consistent with their expectations; P6 reflected that his response from Facebook was “alarmingly unsurprising”. For others, failures to comply fully with the

legislation did come as a surprise: “I am surprised at Google’s unwillingness to provide me with all of the data … they haven’t provided me with all of my data. And that’s not legal.”—P7 Failures to comply led to reflections that “there needs to be more enforcement” (P11), or that data holders are not under sufficient pressure to meet requirements. P6 amended his power evaluation of Facebook to reflect that he perceived them to have total power with respect to his data because the review of the data response had “made it clear which [data] they are prepared to share and which they aren’t”. Likewise, P11 characterised a selective response as an expression of power: “It seems like there’s a lot of derived data about things like purchases and stuff [that I would expect] that just isn’t there. So they’re free to not give me the data. That, to me, suggests they retain an awful lot of power.” —P11

5.4.2.2 A Desire for Understanding

As seen in Table 4, participants shared a common desire to see, know and understand the data that is held about them. At the most basic level, participants hoped to learn whether the data collected and held about them by the organisations exceeded what was required for the running of their services. For example, P11 sought to learn what data is being collected on him when he purchases train tickets: “I’d be interested to understand what data they have […] Is it just the patterns of my spending on trains, or is it a bunch of other stuff that they’re using for advertising to me?” - P11 In particular, participants were interested in data that they would not have been actively aware of. Participants felt that they were aware of the volunteered data (cf. Table 2) that they had deliberately provided; they sought awareness of data that had been collected or derived about them without their knowledge. “The

bit that concerns me is where I don't know that data is being taken by companies. If I'm registering for a library or something, I know [what] data I'm giving to them, but what I don't know is all the other stuff that they're recording" – P9 Participants sought to understand the data in detail and whether it was used to infer further information or to affect decision-making of the data holder. P4 speculated whether data gathered by his smart home lighting might reveal the times at which he typically slept or watched television. P7 reported feeling "weird" about targeted advertisements relating to pregnancy that had most likely been delivered to her based on demographics. This lack of awareness translates to the data holder having greater power "because they're making decisions about things and you don't know how they're making those decisions" (P5). Participants were also curious about the handling of their data and potential sharing between organisations. P4 chose to request data from a company that collated data from credit agencies in order to gain "a picture of what other companies can currently expose". Others wanted to examine the nature of retention and deletion of data, such as P10, talking about dating app Bumble: "Since I found my partner […] I deleted my account and I've been wondering, 'Are they still keeping my data at the back?'" –P10

5.4.2.3 Inadequate Data Responses

While the extent to which hopes were met varied across individuals and data holders (see Table 3), in many cases the desire for greater awareness and understanding was not satisfied. Volunteered data (e.g. basic personal information or user generated content) was most often reported as complete. Participants already expected data holders to have this information and often found it mundane and uninteresting. P6

remarked that Facebook “give you that kind of descriptive boring data which is mainly all publicly available anyway” and that they had omitted “the stuff that I would consider valuable to them”. Frequently, participants commented that they still did not know what data was held, how it was used, or have answers to the questions that motivated their data requests. At the end of the study, when asked his feelings on his data being stored by organisations in general, P11 concluded that: “I still am quite concerned about how much data organisations have, particularly how they link that other data and how data is bought and sold, and I haven’t really got any answers on that.” – P11 In some cases, participants found that returned data was extensive, but excluded context that would give it meaning. P5 received data from a car insurance company that utilises a mobile app for the purpose of generating a driving score and reflected: “I could see the data – it was the score that was weird for me. Like, it doesn’t tell you how it’s calculated.” In addition to the poor quality of data returned, participants often found the process of requesting and accessing the data to be unnecessarily difficult and time-consuming. Four participants (P4, P5, P7 & P11) independently described needing to jump “through hoops” in order to access their data. P10, on hitting a processual barrier, remarked that “I feel like they give you a response that [makes it so] you cannot proceed intentionally”. Participants identified that the painful and tedious processes they had experienced would be off-putting to many. P1 suggested that, without our automated generation of access request wording, it would be “a lot harder to get meaningful data out”. P7 also attributed successful requests to the guidance through the process, and that “even though I did the process correctly, I still didn’t get that much back”. Asked about what she found surprising about one of her requests, P5 responded “how difficult it was just to

get my data, and the fact that I had to ask them about six different times". However, not all of the data requests involved such painful processes: "Some companies make it dead easy to get, but then the data is not massively useful. Other companies make it easy to get, and it's quite useful. Other companies make it a pain in the neck to get it." - P11 Although we found cases of positive experiences, common issues with the process and the responses contributed to frustration and disappointment with the overall effectiveness of the data requests. P10 felt that "frankly, [GDPR] doesn't have as much influence as I expected". Similarly, P1 commented that: "It's kind of disappointing because I would have hoped that this process would have levelled the user power versus the organisation power in a way that holds them accountable and [it doesn't] seem to be doing that" – P1 Despite misgivings and feeling "in the dark" (P4), several participants found the process thought-provoking and report feeling more aware about their data sharing practices and settings. In some cases, this led to intentions to continue with further data requests, such as P4, who found that the process "got me thinking about, like what other things could I try, and what other sources of personal data are there". P8 reflected that "it's a skill and a kind of knowledge that I think everyone should [have]. I don't think it's normal that I felt so clueless". Others commented more directly on the value of understanding GDPR itself, such as P7, who reported gaining "insights into how big companies are actually handling these requests". In the case of P2, this new perspective on GDPR is framed primarily as a lowered expectation of what it can achieve: "[I] think the exercise was useful in the sense that I understand what a GDPR request can do and what it cannot do. And there's a lot it cannot do. And I think it might seem that it gives you a lot of power, but

really, it doesn't." – P2

5.4.3 Theme 2: People Struggle to Understand, Use and Relate to Their Data

5.4.3.1 Data Formats and Usability

Many of the other hopes, uses and plans identified by participants prior to the data requests related to the value that might be found within the data itself (see Table 5). Consistent with personal informatics literature [70], participants anticipated reflecting on and seeking insights in personal data, as well as putting data to practical uses such as budgeting, using it as an aid for remembering and archiving, or using the data for creative or fun purposes. In some cases, potential uses were difficult to predict, as there was some uncertainty about what data would be returned and what level of detail there would be. For example, P5 imagined creatively combining different datasets, but that this would depend on the data format returned. P4 was optimistic about building visualisations of Google data in this way, but uncertain on the detail: "I think … you could do some interesting mashups, but I don't really know what with until I've got the data. It depends on the data; I'm sure there could be some cool uses of it" – P4 Once data was received, participants struggled to interpret and understand it to a sufficient extent to be able to identify the useful data or meaningful information they had hoped for. Returned data formats and response structure were extremely varied, as was the degree to which the data was accompanied by explanations, keys or summaries. The inconsistency in responses was noticed by P1, who expressed that "it would be nice if these companies had a standardised model of how this information is presented to people, so it can be easily understood." However, different recipients with different goals

and technical proficiency would need data in different structures and formats: “They have provided it in formats where I can see that, if I were a developer, I could do things with it, […] but if I was not that sort of person, it might be quite difficult to understand” – P7 P10, who received a response in JSON format, was critical of this, because “for normal people who don’t understand programming, I feel it’s just, there’s no use at all”. A further barrier was the inclusion of information that is meaningful to the data holder but requires further explanation, such as the inclusion of a screenshot of an internal system in one of P11’s responses that was “completely non-understandable”. He went on to question what it can and should mean for data to be usable, as “for me, as a technical person, having a CSV of data is quite useful, potentially, but actually what can I do with that if it’s Tesco’s internal systems data?”. Similarly, P4 reflected that some of the received data “could be valuable if you knew what the hell [was] in there”. Participants also commented on the absence of machine-readable data, as in the case of P4, whose response from his Internet Service Provider included a Microsoft Word document with pasted images from an internal portal. In evaluating its usability and usefulness, he identified that “it depends on what you want to get out of it, really. If you want to view the data they have about you, it’s quite usable. If you want to do something automated, then it’s not.” In some cases, data was considered to be usable, but the lack of explanation or guidance required in order to identify and access the value in their data was problematic. In the words of P5: “They did give me the data, but not how it fitted together. It’s like being given the bricks to a house, and then they’re like ‘Here’s your house’. It doesn’t really mean anything when it’s just bricks, if you don’t know how to put it together.” – P5

5.4.3.2 The Search for Meaning and Value in Data

Problems understanding and extracting value from data were exacerbated when large quantities of data were delivered. P1 compared the variation in quantities across responses as "either like death by thirst or death by drowning – in this case it would be better to drown, but still not ideal". Large quantities of data were harder to digest when presented using obscure formatting and proprietary codes. For P4, when examining data from Philips Hue, there was "just so much of it that it's impossible to know … you'd have to spend a few hours going through this and being like, 'OK, what does that line mean, and that symbol, and that code?'". Participants identified a need for summaries of their data and breakdowns of the data that had been returned. P1, for example, found that one data response was "almost too much […] for a normal person to be able to process and understand what's in there. It could do with a document detailing, like, 'this is what is in here'." Some participants argued that the data received was often not understandable or useable without tools that are designed to simplify or visualise that data "for a non-techie person" (P11). For example, P7 found some of her JSON data more understandable once it had been interpreted via jsonlint.com, an online formatter suggested by the researcher. P2 reasoned that data holders are using certain tools to understand and analyse data "and they're not just looking at a JSON file, so I would like to have the same visualisation [as them]". Some participants did identify parts of their data that were meaningful, useful or otherwise significant. Data that spanned a period of time was considered particularly meaningful. Such historical data was particularly sought after as a means of remembering, with data potentially serving as a "window into your past" (P11). P5 saw value in perusing music-listening

data “just because it’s cool to look back on stuff that you’ve done and you don’t necessarily distinctly remember it”. P6 reflected on some of his data as “a kind of personal history that has been quantified and sort of datafied”. For him, the value around such data is in the small details that form part of the context of certain life events: “I would like to […] build a picture, not just like, ‘I remember going to Reykjavik’, but if there’s other data around that time [I could] sort of paint a biography of myself” – P6 The length of time covered by a dataset also affords individuals with additional capabilities, such as the ability to capture trends and changes over time. The increased value of data covering a larger proportion of one’s lifetime was recognised by P11 when selecting organisations to request data from: “I would actually be interested in last.fm, partly because the data goes back to 2008 … Spotify only goes back about four or five years and not everything I listen to is on Spotify” – P11

5.4.3.3 The Practicality of Using–or Deleting–Your Data

Several participants intended to use data collected over time to better understand themselves and their habits. For example, P2 hoped that his data response would contain, or allow the production of, breakdowns and charts that would help him to learn about his food shopping habits. P10 was able to access details of the accumulation of her spending on micro-transactions on the mobile game Pokémon Go that had not been available to her through the interface of the app itself. P11 had hoped to be able to use train-ticket purchase data to see what he could derive about his journeys in terms of geography, cost, journey length, etc. Participants also considered the increased value of historical data from the perspective of the data holder. For example, P11 saw

mundane data collected over a long period as a liability: “10 years of worth of shopping records … how much would that be worth to a health insurance company, and would they succumb to the temptation to sell that on?”. P10, a Chinese citizen, preferred that data not be retained for long periods, as “in China, [there is a trend] that as soon as someone becomes famous, people begin digging [through] all their past experiences.” Preference for short data retention appeared as a recurring notion. With data holders collecting and retaining data for specified purposes relating to the delivery of services, participants questioned data keeping practices. For example P11 identified that “the thing that kind of concerns me about that is that I haven’t used Tesco online for years, like at least four or five years I think, so why are they hanging on to my IP address from five years ago?”. Most participants described the ability to delete or enforce the deletion of their data as having control over it. In some cases, participants indicated their existing intention to have data deleted, with GDPR expected to play a role in the enforcement or verification of that deletion – particularly in the case of P10, who wished to ensure the permanent deletion of her Bumble and Instagram accounts. In others, deletion was seen as an important option, especially when the holding of sensitive data was considered a liability that was tolerated only in return for an actively-used service: “I now use a different one, but I used, for about a year, an app to track my menstrual cycle. [It was my] main contraception method, so that’s things that this company probably has. Now that I’m not using it any more, I don’t know if they delete the things or not” – P8 Participants foresaw potential uses of returned data to protect their data interests, including checking the accuracy, security and breadth of collected data to flag unforeseen concerns. Many participants hoped to make changes in data habits, privacy settings and choice of service

provider following their anticipated learning from making a GDPR request, that might lead to an increased sense of individual safety and data control: “I want to understand how much they’re keeping. And what they’re doing with it. I’m hoping that by knowing that, I might change my behaviour about all the data I accidentally create.” – P7 However, without better data legibility and explanations, or clear pathways to deleting data, the ability to make such choices was hindered. E.g., P7 remarked: “I guess that’s one of … my criticisms of GDPR in general - that although I can understand what data a company holds about me, there’s no obligation for them to tell me what they are doing with it.. And sometimes I think my willingness to give a company data might be quite intrinsically linked with what they’re gonna do with it.” – P7

5.4.4 Theme 3: Poor GDPR Handling Can Damage the Fragile Trust Relationship

5.4.4.1 Power and Enforced Trust Through Data Holding

A significant proportion of our discussions explored participants’ views on their relationships with providers. We found participants uncomfortable about data collection, especially because of a sense (cf. section 5.1) of being in the dark about current data practices: “I’m curious… I wonder what they’ve got on me. […] If it’s anything other than the barest minimum that is necessary for them to do their job […] then I get creeped out by that.” – P11 All 11 participants expressed the idea that the sacrifice of data is something that they have grown to tolerate in exchange for some benefit. P6 tolerates data collection by travel agents because “they might help me pick a better deal next year.” P11 said he was happy for Tesco to collect data in order to “profile me to try to sell me more

cheese, fine, whatever," though expressed caution that he doesn't "know what else they're doing with it," and more generally was "deeply concerned" about unseen data trading. The benefit can be convenience too; P10 had logged into Pokémon Go with her Facebook account [implying data collection by Facebook] "because it's much easier". This uneasy trade-off surfaced most often in the context of recommendations; generally, participants valued data-based suggestions provided they were "relevant" (P1, P8) and not too "intrusive" (P1, P6). P8 said that relevant music recommendations were "very useful" but found Amazon shopping recommendations "very scary" because "I don't want to see that I'm predictable" and felt that "if someone out there knows [what I want] before you [it's] like taking agency away from me." Participants felt most uneasy about the amount of "intimate" (P1,P2) data that providers collect: "I don't know how comfortable I am with Facebook having as much information as they do about my social circles." – P1 P2 said he feels "quite vulnerable" that his Google search terms "say pretty much everything you have done… the most intimate things you were thinking about". It is clear that data sacrifice is only tolerable up to certain limits: P10 said of Niantic that "as long as they don't sell where I live or my daily routine, I'm fine" and this motivated her to want to check how her data was being used. However, she was critical of their use of in-game benefits as leverage for continued access to users' location data, as "they pressure you into that [and] you don't want to lose out". Multiple participants spoke of such pressure to share data in order to access services or features, and the sense of having no choice. P11 singled out ISPs as having the potential to track everything their customers look at online, noting that "I don't think you've got much choice about that." While data holders typically acquire permission for their collection and

processing of data, P2 felt that giving permission for data collection is “not granular enough”, and in P11’s view “it’s not a negotiation at all, it’s all or nothing.” Accordingly, participants reported feeling resigned about personal data collection: “I feel like it’s inevitable that if you want to access their services at all, in any normal kind of way, that you automatically have to give them your data.” – P7 Participants widely equated the holding of data as a source of power: P7 felt that to have control you need choice and said that “when I think about other people having my data […] the control isn’t sitting with me.” The ability of data holders to limit access to one’s data is also viewed as an indicator of their power: P1 said that “If you’re not getting what you perceive to be yours back in completion then you’re not in control of your own data and you have fairly little power over it.” When asked to define power in the context of data, P8’s comments echoed prior research describing data as a proxy for direct involvement [18]: “For me to have power over my data, I think is a fair and normal thing. But for a company to have power over [my] data means that it’s basically a proxy to have power over me.” The notion of power through data was reflected in participants’ evaluations of power balance: in 69% of relationships participants felt that the data holder had more power than them (rising to 74% after GDPR), whereas in only 17% of cases (unchanged by GDPR) did participants feel they had more power. Participants identified a range of concerns relating to an association of mass data collection with power. P1 noted that companies that “know a lot about everyone will inherently be able to have power either through persuasion or manipulation”. In P6’s view, Facebook used their knowledge of their users’ friendships and relationships to “hook your attention” and prevent users deactivating accounts in a manner that was “disingenuous”. Participants also felt that

some data holders held so much data that it had begun to resemble surveillance, such as in the case of P1, who used “an absurd amount of [Google’s] services” and reflected that “if I’m driving somewhere, I’ve got Google Maps open, so they know exactly where I’m going, they know how fast I’m going, they know what I’m listening to while I’m driving”. Participants feared this kind of deep knowledge of users could be “used against” them (P2). P11 felt that Apple had enough data to “screw me over”, and P5 considered her car insurer Direct Line to be able to use her data to “judge” her, and that “it’s not like I can contest the data and say ‘Actually, no, I disagree’.” In a more extreme illustration, P10 shared her fears that data collected by WeChat and Weibu (the Chinese equivalents of Facebook Messenger and Twitter) would be at risk of abuse by the Chinese government. Ultimately, participants felt that their data was “revealing” (P2,P3,P11) a lot of information about them, and so their only real option to maintain their privacy was to prevent data collection in the first place by not using that service at all (P1,P2,P3,P7,P10,P11).

5.4.4.2 Accountability and Perceptions of Data Holders

Participants said they sought knowledge on how their data is being collected, handled and used (as described in section 5.1) in order to better inform their choice of providers. In learning about how different companies operated with respect to their data, participants sought accountability and whether or not they should make more effort to better manage privacy settings or change behaviours in order to limit data collection. At the beginning of the study, participant opinions about data practices were often based on general factors such as reputation, size or business model. For example, participants who chose Apple as one of the organisations to request data

from (P2, P10, P11) each reported firm pre-existing expectations. P2 described being “more at ease” with Apple, on the basis that their business model was focussed on hardware, than with Google, who were perceived as “making money through data”. He also noted that they were “positioning themselves as a defender of privacy rights”, a point echoed by P11, who was “curious to find out if their marketing claims match their reality around privacy” and wondered if the GDPR request might make him “reassess certain choices”. P10 reported an extreme decrease in trust, not due to the GDPR request, but to a documentary that she had watched between interviews that had caused her to become suspicious of their control over her hardware. Here, GDPR represented an opportunity to compare data expectations with reality. While Apple seemed mainly to benefit from existing trust in relation to their data practices, participants had concerns about other organisations. P6 found that Facebook had “in every shape or form, shown themselves not to be trusted”, an opinion that he supported by referencing “high profile news stories where they have done unscrupulous things and are very willing to just hand over data”. Similarly, P9 reported feeling “slightly dubious” about Amazon due to things that she “had read in the press and about their ethics that may or may not be true, and just the size of them … and just the level of data, as well”. Participants were suspicious of businesses where there was a lack of clarity with respect to how they made money, while those companies that offered a paid service were considered more trustworthy. For example, P8 reported trusting Natural Cycles, stating that “one of the main things was there [are] no ads. It’s a paid service, so there’s no, like, ‘you don’t have to pay but we use your data … to make money’”. Participants also expected good data practices from companies that had made a positive impression on them in unrelated ways. P2

outlined that “when I like the company already, I’m more willing to give them my data”. This was often influenced by the perceived quality of services and software. E.g. P1 found that “in the same way that Amazon is quite janky, Google feels fairly polished and so I trust them more.” Google appeared to benefit from the same effect with other participants, who felt that trust was earned through the provision of valuable services, as with P4, who summarised that “the amount I trust [Google] is in line with the utility I get from them,”. P8, feeling comforted by Natural Cycles’ payment model, also felt encouraged by a sense of shared values: “This is woman-empowerment-orientated, so in that sense I think I do put my trust there as well.” –P8

5.4.4.3 Changed Perspectives Through Scrutiny

Participants’ evaluations of trust in data holders showed a tendency to diminish over the course of the data request process, with some distrust arising following reviews of privacy policy and some following reviews of the data responses (see Figures 5 and 6). In some cases, this could be attributed to an increased awareness of data collection practices, as with P5’s decrease of trust in Spotify after examining their privacy policy “because they shouldn’t need to know that much about me, they should just give me music”. However, it is notable that there does not appear to have been a corresponding reduction in perceived power: “They’ve not given me everything back that I thought they’d be collecting, which makes me trust them less. So power-wise, I don’t think it’s changed, but trust, I think it has.” – P1 As identified by P1, the reasons associated with his lowering of trust scores related not to the data itself, but to the perceived partial or full non-compliance (cf. section 5.1). After receiving her data response from Spotify, P5 lowered her

evaluation of trust further still “because they didn’t say anything about what they’re doing with my data or where it’s going”. P2 reduced his trust score for AirBNB “because of the way they’ve handled [the data request], and the way they’ve made it hard for me to read the data”. Similarly, P7 downgraded her score for LinkedIn “because I feel like they have my data and [they’ve] not bothered to find my data, and that makes me feel like I shouldn’t trust them quite as much”. P8 lowered her trust score for Natural Cycles “because I think it’s hard to get any sensitive data, and it’s not really made clear what they’re using it for”. It appears that some companies have lost users’ trust through a lack of transparency when responding to data requests. In the words of P4, “If someone’s not completely open with you, then you’re like, well ‘What are you hiding?’, which means you trust them less.” “I think the lack of transparency in a lot of these processes has not helped, you know, if Tesco had […] plain English processes for getting the data and you’ve got the data in a plain English way, that would do a lot to bolster trust.” —P11 We saw positive impacts of transparency for data holders in those cases where trust has remained the same or even slightly increased over the process of our study. For example, P5 reflected that her initial view on Instagram may have been “a little harsh” and that she “actually really liked what they sent … in comparison to the three others, I was genuinely, I opened Instagram’s one and I was like ‘this is really cool.’” P10 was very impressed with the response from Niantic and indicated that she trusted them very highly “because they replied really fast, the data provided is very detailed, and their attitude towards this whole issue is very positive,” concluding that they are “a really nice company” and even indicating an increased willingness to spend money on their product. P6 trusted Sunderland AFC

because “they were really kind of upfront and … I got the data from them first, […] no messing about, the format they gave me just made sense.” In these comments, we see an indication that, although the data requests often did not live up to the hopes of the participants, positively engaging with the process was influential and did affect the outlook of our participants. In particular, close attention was paid to the willingness of companies to be transparent and forthcoming, with GDPR representing an opportunity to test organisations on their data practices and assess their integrity and competence as holders of their data.

5.5 Discussion

Our study examined GDPR effectiveness in gaining access and control over one’s personal data. Our participants’ experiences support the established power imbalance (cf. 2.1) and suggest GDPR largely fails to empower individuals: both objectively (to the extent possible by our limited sample), in that most companies do not comply fully, and subjectively, in that returned data was often difficult to understand, impractical for use, and raised new questions and concerns. We also found indications that swift, transparent, and easy-to-use GDPR procedures can positively impact the customer perception of an organisation. In light of these findings, we offer proposals on how the personal data landscape might be redesigned through policy (6.1) and business practice (6.2), and how individual action can have important impact too (6.3):

5.5.1 Policy Implications: Compliance, Quality and Ongoing Access

Despite significant and obvious investment in dashboards, processes and bespoke data package production, our findings

indicate that inadequate compliance with the GDPR is common. Generalisability of our findings is limited by the small number of participants, but they are consistent with literature: we found issues with completeness and compliance – first reported within the GDPR's first year [9] – to persist. However, our focus was on the effectiveness and experience of engaging with GDPR procedures from the individual's perspective. Our participants' experience was overwhelmingly one of disappointment and frustration, with their hopes rarely met. They found that data holders often did not engage meaningfully with the process, and that the responses typically excluded or obscured data that could have provided them with the insights into their data privacy and the organisation's data practices that they sought. Evaluations of perceived power compared to data holders largely remained the same or worsened after accessing data through GDPR, and confidence in the capabilities of the legislation to shift the balance of power was lacking. The process was perceived by some as a "box-ticking exercise" that was both frustrating and time-consuming and did not ultimately help them. Even though in 7% of cases participants did feel empowered by the GDPR, all participants receiving data were in practice left with additional time-consuming and sometimes technically-skilled work to take advantage of or interpret their returned data. This suggests that to improve, policymakers need to make changes towards:

1. Better Compliance Through Enforcement of Complaints. At present, enforcement of the GDPR is uneven; each country has its own DPA (for example in the UK, this is the Information Commissioner's Office or ICO) and complaints are rarely pursued for individual cases. Instead, cases are processed by specific DPAs in a form similar to a class action lawsuit. This means that individuals have little

impact when they do raise a complaint, and many GDPR complaints “become lost or resulted in lengthy delays” [21], or may even be erroneously dropped [72]. Until individuals have a clear and effective means to issue complaints [11] that result in enforcement action (or a clear threat of it), it is likely that individuals will continue to have little recourse other than to repeat the request and hope similarly dissatisfied individuals will act on their behalf. Data holders must be held to account when they do not deliver the full set of data that they report possessing, or when they fail to do so within the legally obligated time frame.

2. Policies to Enforce Better Quality Responses. Many participants received data in frustrating formats, including screenshots, printouts or files that were too technical or littered with acronyms. Data was provided in formats too technical to understand, or not technical enough to be usable (see 5.2), showing a demand for both human-readable information summaries and machine-readable data files, where most providers typically provide only one or the other. Policymakers could provide suggested data formats or even propose new standards; this would help data portability, improve effectiveness [44] and legibility [80], can reduce costs through common tooling and catalyse the building of tools to interpret and understand data. Such standards are emerging [78] as they are a technological necessity for data unification, but lack adoption.
3. Policies to Enforce Data Access as Ongoing Support, not One-Time Delivery. We believe a radical redesign of policy is needed to give people the practical outcomes they desire and, according to the GDPR itself, deserve. Data access needs to be seen as more than “the delivery of data files”. People need understanding of their data and of

its handling, and this is the measure by which compliance should be assessed. The explanations GDPR mandates are not forthcoming; of the 119 hopes expressed by participants (see Table 5), 70 (59%) related to acquiring greater understanding of data practices. 38 (54%) of these were unmet, and a further 15 (21%) were only partially met. By mandating data holders to support individuals with not just the delivery of data, but assistance to understand that data, policies could become more impactful, not least because such understanding is critical to inform judgements around consent, loyalty and compliance.

5.5.2 Implications for Business Data Practices: Earn Trust through Opening Up Data

While this study, and the GDPR itself, might seem adversarial to data holders given the goal to reduce their power by imposing new procedures, our findings emphasise the role of personal data in user relations. Data holders are likely aware of the paramount role of personal data in decision-making, but may not be aware of users' perceptions about this. Our findings suggest that failure to satisfy users who are concerned about the collection and usage of their personal data risks harms to user trust and confidence. In turn, however, this presents opportunities to use the mechanisms of the GDPR for user loyalty and building better relations. In 52% of cases, following our process of examining privacy policies and engaging in GDPR data requests resulted in a decrease in reported trust in the data holder. While such impacts may for now be minimal, as only a small proportion of users read privacy policies [92] and—we can assume—an even smaller number conduct GDPR requests, this is likely to change as issues around data privacy and trust continue to take centre

stage in global geopolitics [98,107]. Furthermore, the growing number of businesses focused on “getting your data” or “taking control” [25,40,97,116–118] suggest demand for data access is growing. , this is likely to change as issues around data privacy and trust continue to take centre stage in global geopolitics [98,107]. Furthermore the growing number of businesses focused on “getting your data” or “taking control” [25,40,97,116–118] suggest demand for data access is growing. Our findings offer three positive takeaways for data holders: 1) Data transparency is an opportunity to increase user loyalty and trust. GDPR’s basic rights provide a starting point for delivering practical data transparency that will allow organisations to demonstrate that they are deserving of trust. By responding clearly and engaging openly and helpfully with GDPR data requests, organisations can demonstrate consistency between their privacy policy and their actions and demystify to their users the role that data holds in their business model. Research has shown that explanations can “ease humans’ interactions with technology […], help users understand a system’s function, justify system results, and increase their trust” [42]. In 14% of cases, our participants felt more trusting of the service brand as a result of their GDPR experience (sometimes even displacing prior apprehensiveness or distrust), citing reasons such as speedy, hassle-free responses, clear and understandable data, providers being upfront and open with data, and staff who exhibited a positive attitude to the request. 2) Data transparency is an opportunity for improved and re-imagined “user data relations”. Beyond the opportunity to improve trust, the mechanisms of data transparency suggested by the GDPR provide individuals with new capabilities for data curation and involvement. By offering users the ability to engage in empowering data interactions, data holders have the

opportunity to improve engagement with their organisation and their services. If organisations view user data as a shared resource to be curated and co-owned by the individuals that contributed it, there may be correspondingly shared benefits: for the individual, a sense of agency, influence and negotiability [80]; and for the service provider, an incentive for users to generate and share more data, an increased likelihood of users correcting inaccurate data, and more reliable and human-centric forms of ongoing consent closer to dynamic consent [61] than today's ineffective models of informed consent [73]. 3) New user demands indicate untapped business opportunities. As the 500-member-strong MyData Global organization [82] shows, there is growing demand for personal data empowerment. People's personal data is splintered and trapped [1,16], and they cannot correlate data from different sources in order to reflect upon it, gain insights, and set goals [70]. Due to commercial motivations, service providers generally deliver capabilities within a closed silo, not at the level of one's wider environment [2]. To be better empowered the individual could be the point of integration, the centre of their own Personal Data Ecosystem (PDE) [81]. Life-level capabilities [17] and the opportunities that well-designed and well-regulated GDPR-type regulations promise in this regard have not yet been exploited. Thorough, complete and timely data access in standard formats, as mentioned above, will be critical to enabling this vision. Growing companies such as CitizenMe [119], Digi.Me [38], Mydex [125], ethi [58], HestiaLabs [34], udaptor [97] and exist.io [120] as well as larger organisations like BBC R&D [13] and Microsoft [75] are already starting to innovate in this space.

5.5.3 Individual Action: Becoming Aware of the

Value and Power of Data

While participants experienced disappointment and frustration in their GDPR journeys, all participants gained new understandings; if not always of their data itself, at least of their target companies' approach to data access requests. This new knowledge was sufficient to re-affirm or challenge existing attitudes or inform judgements –P1, for example, left Facebook after the study. Even an attempt to access data can be educational, and even a cursory look at a provider's 'What data do we collect' privacy policy section can provide pause for thought. Today, individuals remain largely in the dark about the collection, use and sharing of their data through a combination of perceived complexity and effort combined with a lack of clear benefits. Table 4, alongside the increased control and insight promised by the PDE movement and platforms linked in 6.1 and 6.2 above, provide a glimpse of what the future may hold: a world where individuals take more control of their data and gain actionable self-insights. We infer three key messages for individuals: 1) Your data is used to represent you and define your user experience. We hand over our data in exchange for access to services, but providers then use it (usually in aggregate) e.g. to inform product design or decide what content you see. This 'innocent' handover of data is in fact giving providers the means by which we are treated and – at times – controlled. Recognizing the crucial role of data (and our limited influence over it) is the first step to pursuing greater agency and control. 2) Your data contains meaningful and valuable data about your life. Data, as our participants found, is dry and technical, but they all sought meaning and value within it (cf. 5.2.2). Within data lies potentially rich information about one's life and past activity – some of which can even be inaccessible through any other means. This highlights both a

risk (that others might gain this insight) and a potential benefit (that we could access this insight ourselves). In this context, data deletion without keeping a copy may be inadvisable. To access the value in data, individuals will need to demand data standards, better access and control mechanisms and insight tools. 3) Self-education and awareness enable accountability and informed choices. Our findings highlight a lack of knowledge. Transparency is critical to judging ‘to what extent the bargain is fair’ [66]. It is not always delivered, but GDPR makes it your right; a right that cannot be fully refused. Through challenging poor GDPR responses and demanding better information, individuals can have impact. Providers are ultimately motivated by user demand—one of the reasons download dashboards exist. Through the public pressure of negative attention, companies can be motivated to improve data access [33]. With patience, GDPR rights can be exploited to force small changes.

5.6 Summation

We set out to evaluate the human experience of GDPR for service users; through our longitudinal qualitative investigation of attitudes to data-centric services before, during and after conducting data access requests, we uncover shortcomings in providers’ GDPR approaches that result in an unsatisfactory experience for users. When encouraged to draw conclusions on the basis of providers’ own promises, individuals’ legal rights, and their own hopes (see Table 4), participants, initially lacking awareness of held data and its uses, gained some insights, but were generally still ‘in the dark’. We found that providers can seriously damage brand loyalty and trust if comprehensive and well-explained data is not returned in a supportive and open manner (see 5.3). We highlight serious

problems with compliance (see 5.1): Participants received data that was incomplete, impractical for use, and they failed to acquire desired explanations. By its own aim to enhance individuals' rights and control, the GDPR does not succeed. Participants continued to feel a lack of agency and choice, were largely unable to pursue goals such as data checking, correction or deletion, and their perceived sense of power within the provider relationship was largely unchanged by the experience. Nor does the GDPR allow individuals to adequately pursue their own goals related to accountability, self-reflection or creative data exploration (see 5.2). Individuals cannot be given power over their data through designing better Human-Data Interaction interfaces alone, but only through redesigned policies and business strategies that take into account the sociotechnical context [12,17]. From a policy design standpoint, action must be taken to improve both compliance and quality of GDPR responses, but better still would be to provide individuals a more effective and ongoing two-way window into their data (cf. 6.1); data access is a lifelong pursuit. The risk of reputational damage should motivate data holders to engage meaningfully with data access requests; such risk can be averted by redesigning both interfaces and processes to approach data access experiences as an opportunity to build trust and loyalty, perhaps even through establishing progressive co-operative data stewardship relationships (cf. 6.2). While the GDPR experience is often disappointing and frustrating, it can provide insights that help individuals to challenge their assumptions, re-evaluate choices, and in some rare cases, feel empowered to act upon their data. Wider assertion of GDPR rights could demonstrate a desire for data holders to be transparent; without such visible demand, little may change (see 6.3).

Bibliography