Report on *Living With Data* Interviews & Focus Groups

July 2022

Hannah Ditchfield, Helen Kennedy, Susan Oman, Lulu Pinney, Jo Bates, Monika Fratczak, Itzelle Medina-Perea, Mark Taylor

Contents

A. INTRODUCTION
1. About *Living With Data (LWD)*
2. About *LWD* interviews and focus groups
3. About participants and recruitment

B. FINDINGS
4. Fairness and inequalities: what 'good' data uses look like
4.1. Evaluating the fairness of data uses
4.2. Inequalities and perceptions of data uses
5. How people understand data uses
5.1. The role of imagining in understandings of data uses
5.2. The role of everyday experiences in understandings of data uses
6. Perceptions of data uses can change

C. CONCLUSIONS
7. Talking to people about data uses
8. Main findings and recommendations
9. References

Living With Data is funded by The Nuffield Foundation.

A. INTRODUCTION

1. About *Living With Data (LWD)*

Living With Data (LWD) is a research project which aims to understand people's perceptions of how data about them is collected, analysed, shared and used, and how these processes could be improved. We use the term 'data uses' as a short and accessible way of talking to people about these processes. The data at the centre of such processes is often personal data, defined as data 'related to an identified or identifiable person' by the General Data Protection Regulation (or GDPR, European Union regulation about data usage and rights).

This document reports on interviews and focus groups on perceptions and opinions about data uses, undertaken with 112 adults in the UK from November 2020 to September 2021. A summary version of this report can be found in the Resources section of the Living With Data website, along with other publications from the project.

Uses of data can result in harms as well as benefits. Concern about potential harms has led to initiatives which aim to ensure that data works 'for people and society' (the mission of the Ada Lovelace Institute). This concern also motivates our research, as we believe that understanding how people think about data uses can inform efforts to minimise harms.

There has been an increase in studies of public perceptions of data uses in recent years, but there are some gaps in research and understanding, which we aimed to fill with our research. These are:

- There is a lot of debate about whether **fairness** is a useful lens for assessing data uses, but the perspectives of non-expert, ordinary people are missing from these debates. Understanding what kinds of data uses people consider to be **fair** is important, in order for debate about improving data uses to be informed by public views. We write about what we found about fairness in Section 4.
- Research has found that **social inequalities** play a major role in shaping people's *experiences* of data uses. There is less research into whether and how various social inequalities shape *perceptions* of data uses. Research which helps us understand the relationship between social inequalities and perceptions will help us identify whether and how data uses can be improved. We write about what we found about social inequalities in Section 4.
- While there is interest in the **relationship between awareness of data uses and attitudes towards them**, research which explores this relationship is limited. Are people who are more aware of data uses more or less concerned about them? To answer this question, we need a better grasp of what it means to **understand** data uses, how people come to understand them and the roles that **imagining**, **feelings** and **everyday experiences** play in people's understandings of data uses. We discuss what we found about these topics in Section 5.
- A lot of previous research has focused on attitudes to high-profile data practices. We also need to understand people's views of **everyday data uses**, such as those that take place in the public sector, because of the role that they play in shaping everyday lives. In addition, a lot of research explores attitudes towards data uses in general or in the abstract. In order to improve data uses, we need to understand people's perceptions of **specific data uses in specific domains.** Our

research focused on specific public sector data uses, as we describe in the next section.

We addressed these issues by asking two research questions:

1. *What do different people know and feel about specific data-related practices in different domains of everyday life?*
2. *What do fair data practices look like, from non-experts' perspectives?*

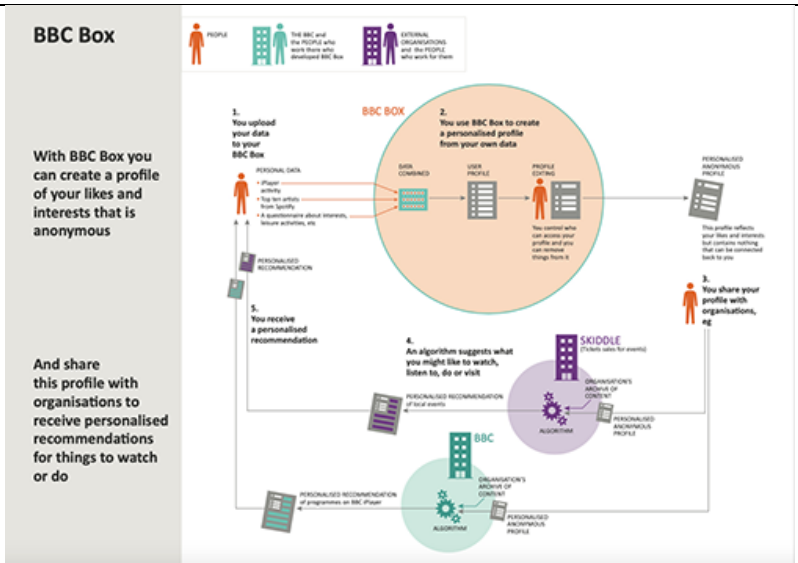
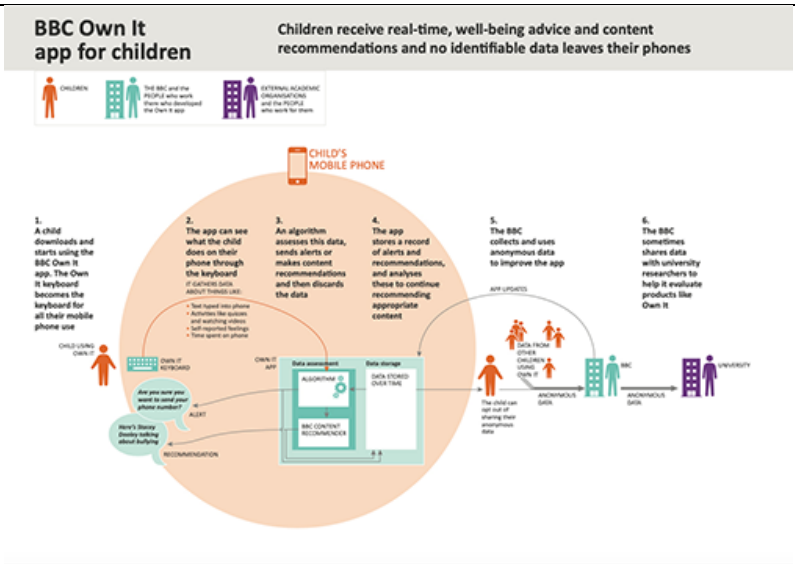
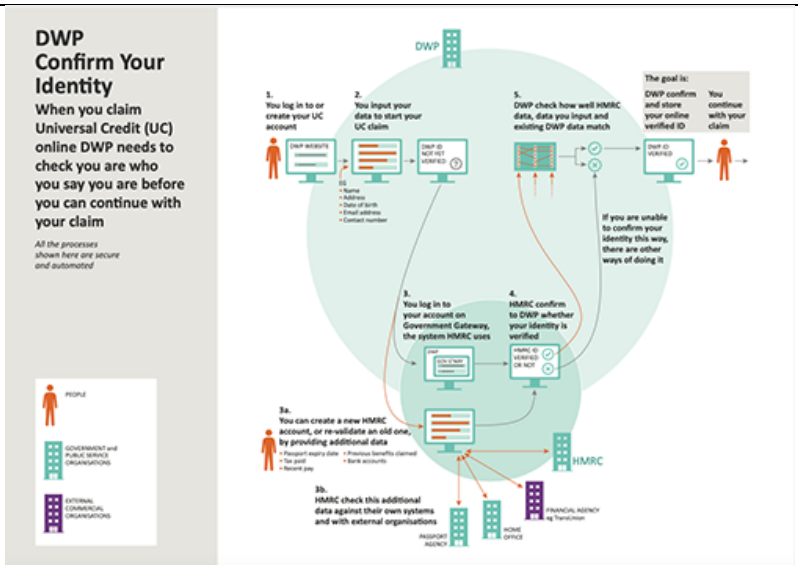
2. About *LWD* interviews and focus groups

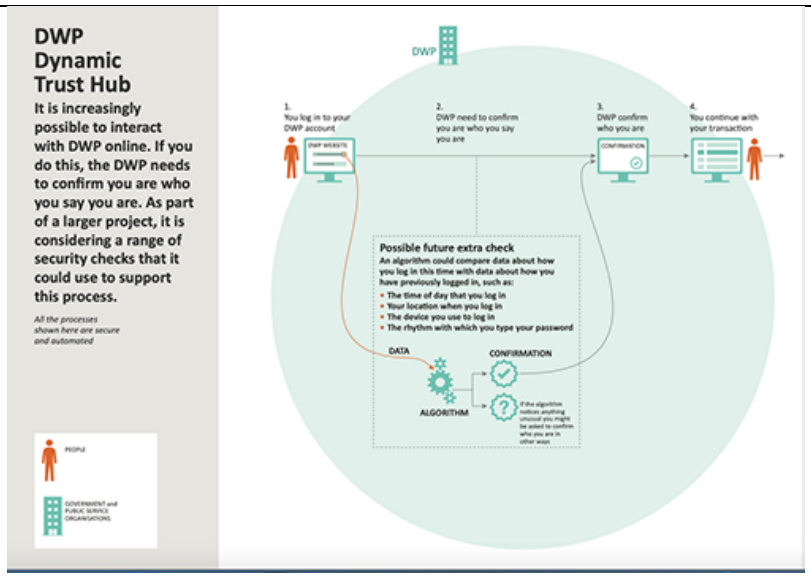
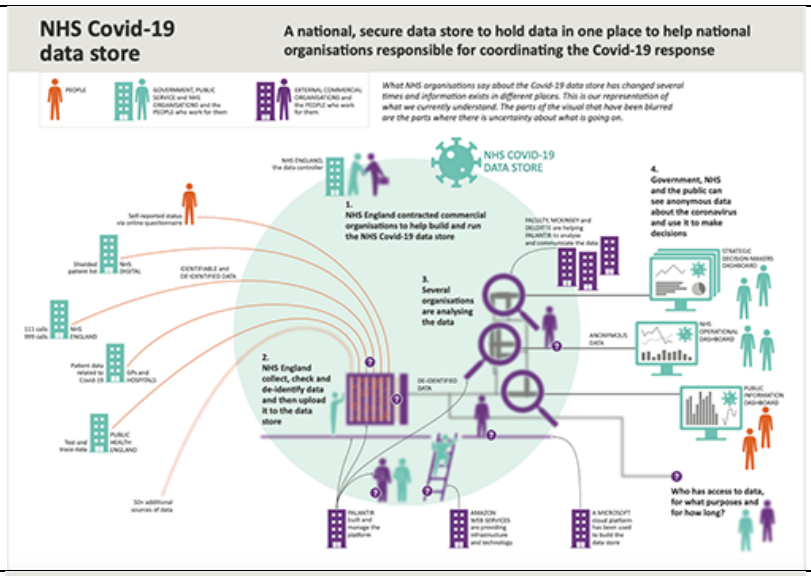
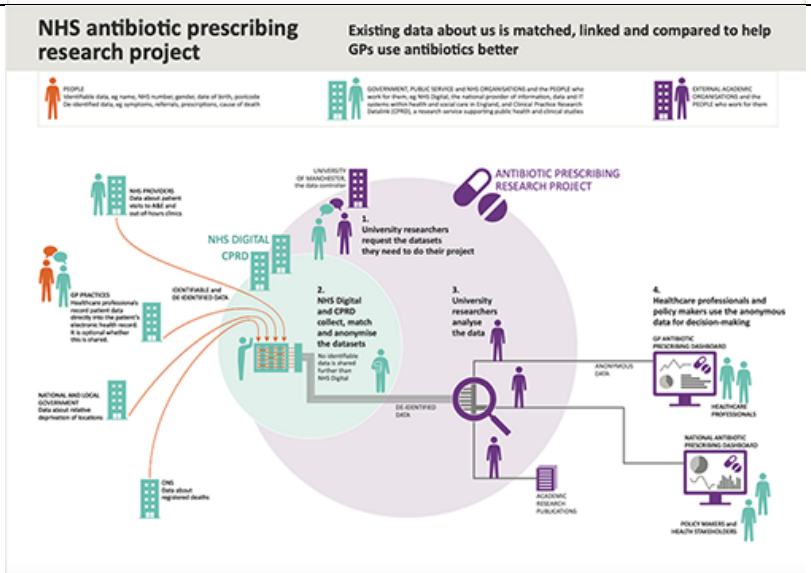
We produced accounts of specific data uses which we presented to focus group and interview participants for discussion, in order to research people's knowledge and feelings about specific data uses. We selected cases from the public sector because public sector data uses shape everyday life experiences, and yet had received less attention than high profile commercial data uses at the time of our research. We identified **welfare, media** and **health** as three domains on which to focus our research because they are core aspects of everyday life. To produce accounts of data uses in the first two domains, we partnered with the government Department for Work and Pensions (DWP) and the British Broadcasting Corporation (BBC). For the latter domain, health, we produced an account of a data use based on information in the public domain, and another account based on research that one of us, Itzelle, had undertaken into National Health Service (NHS) data flows (Medina Perea et al 2019).

To produce accounts of data uses, we deployed some of the techniques used by one of us, Jo, (in Bates et al 2016) to map data flows and frictions, as part of an approach they call Data Journeys. This process involved developing detailed knowledge of each data use, through interviews and textual analysis of partner organisation and publicly available documentation. In the data uses which were the focus of our research, data did not always flow, or go on a journey – indeed, some data uses aimed to limit the movement of personal data in order to give people more control over data about them.

The BBC and DWP data uses were selected by our contacts within these partner organisations. We focused on two BBC experiments with personal control over data, two DWP data uses which both focused on ways of verifying identity online and two diverse health data uses. We produced accounts and visualisations of these data uses, which we used in our research to show and explain data uses to participants. Figure 1 below shows small versions of the visualisations we shared with focus group and interview participants. Full-size visualisations of data uses can also be found on the Data Uses page of our website.

Figure 1: Short descriptions, thumbnail images, benefits and harms of the six data uses which were the focus of our research

<u>BBC Box,</u> a prototype which pulls together data about what users watch or listen to and gives them control over who has access to this data. **Proposed benefit:** people control their own data & who has access to it (2); people receive personalised recommendations (2, 4) **Potential harm:** whether data is secure if all in one place (2); whether individuals have the time & expertise to manage their own data (2); recommendations systems can recommend a narrow range of things, and people using them are not exposed to a range of issues of perspectives. Some people say this is not good for democracy (4)	
<u>BBC Own It,</u> a free app designed to support, help and advise children when they use their phones to chat and explore the online world, without adult supervision. **Proposed benefit:** children receive personalised wellbeing advice & support without identifying data leaving their device (3, 4) **Potential harm:** interventions in children's wellbeing take place without parent/carer knowledge (2); potential for anonymous data to be shared with researchers, which may not be clear in data agreement (3); automated, makes assessments of & recommendations about children's wellbeing, without the involvement of a parent/carer (4)	
<u>Confirm Your Identity,</u> an identity verification process for Universal Credit payments which makes it possible to confirm identity online. **Proposed benefit:** don't need to confirm identity in person/with paper docs (1) **Potential harm:** requires people to have HMRC account, passport, bank account, financial record, so excludes people with complex lives; negative consequences for people whose identities can't be verified (1)	

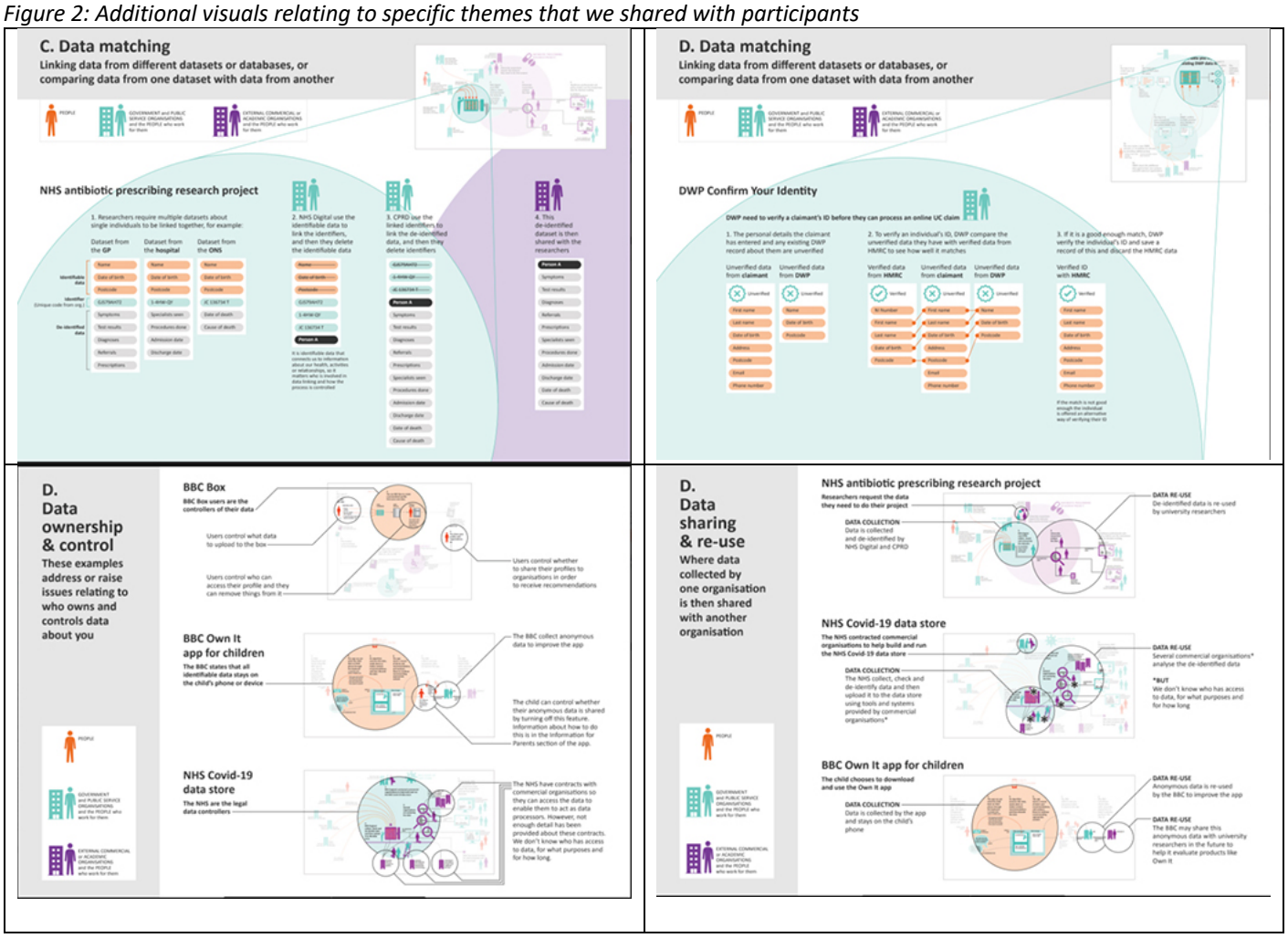
<u>Dynamic Trust Hub</u>, which explored a range of issues to enhance identity verification, including technology integration and possible security checks. **Proposed benefit:** don't need to confirm identity in person/with paper docs & extra layer of security to the identity verification process (4) **Potential harm:** requires people to have HMRC account, passport, bank account, financial record, so excludes people with complex lives; negative consequences for people whose identities can't be verified; inaccuracies in automated checks (eg people may not own their own devices) which may be taken as a proxy for unverifiable identity (4)	
<u>NHS Covid-19 Data Store</u>, a national data store to help organisations responsible for coordinating the Covid-19 response. **Proposed benefit:** data gathered together in one secure place to inform the national response to the pandemic (2, 3) **Potential harm:** lack of clarity about who has access to data & the contractual arrangements with commercial organisations; commercial companies may be able to use data from the data store to develop their technologies and potentially increase their profits (2, 3).	
<u>NHS antibiotic prescribing research project,</u> exploring ways to address the public health crisis of antibiotic resistance**.** **Proposed benefit:** addresses public health crisis of antibiotic resistance (1, 3) **Potential harm:** requires the use of patient data (1); whose data is shared (could it be children's?); who the data is shared with (could it be pharmaceutical companies, who could eventually profit from it?) (3)	

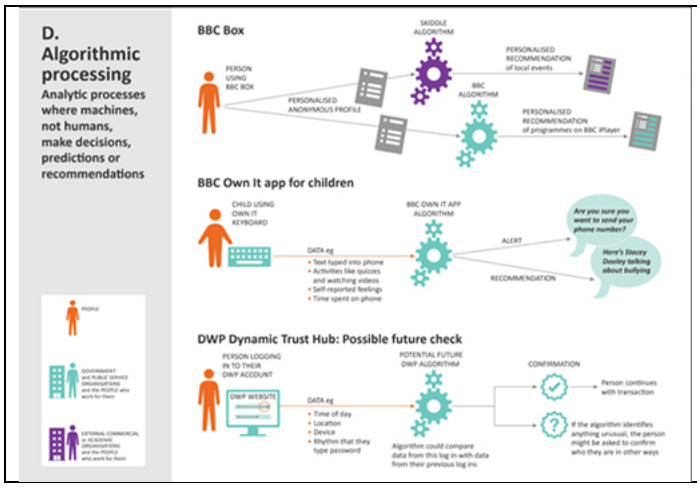
For the interviews and focus groups, we grouped our example data uses into four themes. The themes were:

1. **Data Matching:** this is where organisations match data from different datasets or databases, or compare data from one dataset with data from another. The Data Matching theme included the NHS antibiotic prescribing project and DWP Confirm Your Identity.
2. **Data Ownership and Control:** this theme focused on who owns and controls data about us. It included both BBC data uses, BBC Box and BBC Own It, and the NHS Covid-19 Data Store.
3. **Data Sharing and Re-use:** This theme focused on instances where data collected by one organisation is then shared with another organisation. It included the two NHS data uses, the NHS antibiotic prescribing project and the NHS Covid-19 Data Store, and BBC Own It.
4. **Algorithmic Processing:** this theme focused on analytic processes where machines make predictions which can lead to recommendations to human operators or automated decisions. It included the two BBC data uses, BBC Box and BBC Own It, and DWP Dynamic Trust Hub.

We produced theme-based sets of visualisations, adding visuals that compared data uses included within the theme or more information relating to the theme. Figure 2 below contains the additional visuals which were shown to participants for each of these themes; there are two visuals for the Data Matching theme, and one each for the other themes. Full-size visuals can also be seen online, here.

Figure 2: Additional visuals relating to specific themes that we shared with participants





Each focus group or interview focused on one of the four themes. We asked some general introductory questions about perceptions of data uses, and some common closing questions about perceptions of fairness, extent of understanding, and the importance of the context of the data uses. The majority of each focus group or interview was dedicated to a set of questions about the specific data uses in the chosen theme. The questions that we asked can be found on a web page about our qualitative research.

We used our accounts and visualisations to show and explain data uses to participants, to make details about potentially unfamiliar data processes accessible to them. We did this because we believe that showing and telling people about data uses enables people to develop and express opinions about them, and that we can't expect people to have opinions about data uses if they don't understand them. This is especially true in relation to benefits and harms: people can't assess potential benefits and harms if they don't know what they are. For this reason, we told participants about one proposed benefit and one potential harm or concern for each data use, which sometimes varied according to the theme under discussion. Shortened versions of benefits, harms and concerns can be seen in Figure 1 above, and the longer versions that participants heard can be found online on this web page about our qualitative research.

Providing such information to participants may have encouraged them to think about benefits and harms that they may not otherwise have considered. We believe that it is not possible to produce objective accounts of data uses, because interpretation takes place in the acts of describing and illustrating them, choosing aspects to highlight and deciding what to include. We tried to be accurate and balanced in the ways we described data uses and represented them in visualisations, aiming to produce simple but detailed-enough accounts, working iteratively on our accounts and visualisations, in our team, with our partners and our advisory group.

3. About participants and recruitment

112 people participated in our focus groups and interviews. Because we were interested in the relationship between social inequalities and attitudes to data uses, we recruited demographically diverse participants. Demographic information was gathered via a 'Tell Us About You' survey, which 100 of 112 participants completed prior to the interviews and focus groups. We gathered information about age, ethnicity, gender, sexuality, country of birth, nationality, employment status, household income, education, whether English was a first language, long-term conditions. We also asked participants about the newspapers that they read

and their degree of interest in politics. Participants had the opinion to choose 'prefer not to say' to any of the questions. They could also choose to provide some information in free text fields, about their job title, the ages of children in their households who were under 18, their country of birth and the newspapers they read. We gathered this information to help us interpret the research data that we gathered, because previous research had suggested that these factors may play a role in shaping attitudes.

In this report, we share some of the information about participants that we gathered. Sometimes the information is relevant to the things that they said and the views they expressed. Sometimes we share information to paint a picture of the person we are writing about, or to avoid 'othering' – that is, highlighting only minority characteristics. We describe participants' gender, sexuality, race, nationality or ethnicity, age, occupation, income and whether they have long-term conditions, where we have this information. We don't share all of the information we gathered about each participant to preserve participants' anonymity, something which is especially important when writing about disadvantaged, vulnerable or minority groups (Fox et al 2021). We also use pseudonyms to refer to participants as a means of preserving anonymity.

Our research took place during the Covid pandemic. Lockdown conditions made it difficult to recruit participants from certain disadvantaged communities who may have limited access to the tools and skills needed to engage in virtual meetings. Nonetheless, we managed to recruit a sample of which:

- 44% of participants were Black, Asian or other racialised people;
- 41% had household incomes of less than £19k a year;
- 37% had a long-term condition;
- 20% were LGBTQ+;
- 14% were 65 or older;
- 46% had two or more of these characteristics;
- 15% had none of the above characteristics.

Further demographic data about participants can be found in Table 1 below.

We used definitions from the 2010 Equality Act to identify whether participants had disabilities, asking if they had any physical or mental health conditions or illnesses lasting or that were expected to last 12 months or more. Also following the Equality Act, we use the phrase 'had a long-term condition' to refer to participants with disabilities, although we recognise that conflating health and disability is sometimes challenged by disability activist and advocacy groups.

For the question about country of birth, the options were: UK; other European country; other non-European country; prefer not to say. Some participants chose to tell us their country of birth in free text fields. In response to the question about ethnicity, we presented participants with the 19 categories used in the Census in England, which we then aggregated into: White British; White Other; Black; Asian; Mixed ethnicities/other; prefer not to say. We recognise that there are more ethnicities than those we use in this report. We chose to use these six categories because to use more categories than this would not be productive in our analysis of our 112 participants we aimed to recruit for the focus groups and interviews. In this report, following Sobande and others, we use the term 'Black, Asian and other racialised people' to refer to participants in the Black; Asian; Mixed ethnicities/other categories. We do this because we acknowledge criticisms that the term BAME (= Black, Asian and Minority Ethnic) is reductive (#BAMEOver; see also Katwala 2021; Malek et al 2021). We also acknowledge that there is no one obvious term to use to describe participants who experience racism.

Table 1: Demographic data about participants

Variable	Category	Proportion
Activity/employment	Working full-time	26%
	Working part-time	15%
	Unemployed and looking for work	11%
	Furloughed	2%
	Full-time university student	6%
	Other full-time student	2%
	Retired	10%
	Not in paid work for any other reason	13%
	Other	9%
	Prefer not to say	7%
Total household income	£19,999 or less	42%
	£20,000 to £39,999	20%
	£40,000 to £59,999	9%
	£60,000 to £100,000	9%
	£100,000 or more	2%
	Prefer not to say	18%
Highest education qualification	None / don't know	8%
	Post-16 qualifications (eg GCSEs)	5%
	Post-18 qualifications (eg A-levels)	5%
	Some higher education (eg nursing qualification)	14%
	Undergraduate degree	22%
	Higher degree (eg MA/PhD)	24%
	Other qualification	26%
	Prefer not to say	4%
Age	18-24	8%
	25-34	24%
	35-44	36%
	45-54	7%
	55-64	11%
	65 or older	14%
Country of birth	UK	58%
	Other European country	4%
	Other non-European country	32%
	Prefer not to say	6%
Citizenship	UK	76%
	Other European country	5%
	Other non-European country	17%
	Prefer not to say	2%
English as first language	Yes	61%
	No	39%
Have children under 18	No	64%
	Yes	36%
Long-term condition	Yes	36%
	No	62%
	Prefer not to say	2%
Impact of long-term condition on day to day activities	A lot of impact	30%
	A little impact	43%

	No impact	24%
	Prefer not to say	3%
Type of long-term condition	Mobility	17%
	Lifting, carrying, moving objects	20%
	Manual dexterity	3%
	Continence	3%
	Communication	7%
	Memory, ability to concentrate (to learn and understand)	12%
	Recognising when in physical danger	3%
	Physical coordination	3%
	Other health problems	22%
	Prefer not to say	10%
Ethnic group	White British	48%
	White Other	8%
	Black	9%
	Asian	20%
	Mixed ethnicities / other	13%
	Prefer not to say	2%
Gender	Male	30%
	Female	62%
	Other	4%
	Prefer not to say	4%
Sexuality	Heterosexual / straight	74%
	LGBTQ+	19%
	Prefer not to say	7%

We describe as LGBTQ+ participants who answered 'no' to the question 'is your gender the same as the gender you were assigned at birth' and who described themselves as bisexual, asexual, lesbian/gay, queer or other. We recognise that there will be differences across these groups, but as above, given our relatively small number of participants, we felt that this grouping was most appropriate for our analysis.

Recruitment was conducted in three waves: November 2020 – April 2021; May 2021 – July 2021; and August 2021 – September 2021. In between each wave of recruitment, we reflected on who we had spoken to and identified who we needed to prioritise in the next recruitment wave. These reflective pauses allowed us to see that certain intersections were missing from our sample, such as older Black, Asian and other racialised people, or people on lower incomes who were also not in full time education.

Because of the pandemic, groups of people who normally met up to socialise – for example, by playing sports – were finding other ways to stay in touch, and this informed our recruitment strategies, which included:

Social media pages/accounts: We posted messages inviting people to participate in our research on various social media groups and pages. We attached images to these posts which summarised key information about the project, and we accompanied these images with further details in the caption. We posted on:

- Neighbourhood-focused Facebook groups
- Community-focused Facebook groups (eg groups for Latin Americans, LGBTQ+ groups)

We also sought to recruit via the *Living With Data* Twitter account.

Charities and community organisations: We contacted relevant charities and organisations which worked with or had connections to groups of people to whom we wished to talk. Once contact with a gatekeeper at an organisation was established, the gatekeeper would either organise a focus group for us or circulate details about our research within their community, via email, social media or word of mouth. People interested in talking to us would then contact our research team directly.

Hobby/interest groups: we contacted a large variety of groups that gathered together based on their interests. This included religious, activist and sports groups, and musical or creative classes. As with charities and community organisations, once contact with gatekeepers was established, they either organised a focus group for us or circulated details about our research.

Other: We leafleted targeted homes, such as council-owned tower blocks, once lockdown restrictions were eased, in order to attract participants in areas populated by communities with whom we wished to talk. We used local council data to identify suitable areas. We also used snowballing to recruit participants. In other words, sometimes participants would pass on project information to people within their own networks, who they thought would also be interested in taking part in our research.

Considering inequalities in qualitative online research

Due to the online approach of data collection most of our participants had access to the online technologies needed to take part in the research. 53% of our participants claimed to use the internet for personal use 'several times a day' and 35% 'almost constantly'. 58% were 'very confident' using devices to do things online and 33% somewhat confident.

Aware that some of the participants we wished to recruit may have limited access to devices and the internet, we made provisions to carry out our research in ways which were inclusive, including:

- phone calls were offered to people who preferred them to online video calls;
- print outs of visualisations of data uses, information sheets and consent forms were sent by post to people who preferred this to receiving them via email;
- we used a mix of digital and non-digital recruitment methods to reach groups not online, on email or on social media;
- once pandemic conditions allowed, two face-to-face focus groups were conducted with low income women who otherwise would not be able to take part;
- we set up a research mobile phone number for participants who wished to communicate with us this way or who did not have access to email; this proved popular with some groups who were more comfortable arranging interviews on WhatsApp than via email
- we worked with facilitators to enable people with cognitive or learning disabilities to participate.

B. FINDINGS

4. What 'good' data uses look like: findings about fairness and in/equality

Concern about the potential harms of data uses means there is a lot of interest in ensuring that data 'works for people and society' (Ada Lovelace Institute), as noted in the introduction to this report. But multiple, often contradictory ideas about 'the good' exist, and there is no consensus on how to evaluate or advocate for 'good' data uses. People's understandings of whether, how, when, where and for whom data uses are good are rarely central to research, policy and practice. This limits efforts to ensure that data uses work for people and societies. To understand what better, improved or 'good' data uses look like from non-experts' perspectives, we asked our participants what they considered to be **fair data uses**. In responses to these questions and in our broader conversations with participants, we found that **the relationship between data uses and social inequalities** was an important factor in their reflections about good data uses.

4.1. Evaluating the fairness of data uses

On *Living With Data,* one of our main research questions was: *What do fair data practices look like, from non-experts' perspectives?* In 2018, Verma and Rubin identified at least twenty definitions of algorithmic fairness that had emerged in recent years, within the fields of law, software engineering and AI. In everyday usage, which is what concerns us here, fairness can refer to impartial or just treatment, treatment without discrimination or harm, or what is reasonable, right and just. Despite a lack of consensus about its meaning, researchers exploring perceptions of data uses have found that fairness is a useful concept. In 2015, one of us, Helen (writing with others), argued that a concern for fairness was common among participants in research exploring what European social media users think about social media data mining (Kennedy et al 2015). They argued that participants weighed up a range of factors, including the type of data tracked and gathered, the purpose of the data mining activity, the extent to which social media activity and data were perceived to be public or private, and transparency and consent, asking themselves: is this specific data mining practice fair? They concluded that this concern for fairness could be understood as part of broader concerns about well-being and social justice.

In recent years, a number of initiatives have emerged which aim to promote fairness, along with accountability, transparency and ethics in data-driven systems, AI, automation and machine learning. This move results in part from recognition of the fact that data-driven systems are imbued with the values of their designers, with worrying consequences for people from disadvantaged or minority groups who are underrepresented amongst data systems designers. Fairness, accountability, transparency and ethics initiatives (variously abbreviated to FAT, FATE, FAccT and FACT) bring together researchers and practitioners, including data scientists, legal experts and ethicists, who acknowledge that data-related technologies can introduce bias and discrimination when they are deployed, and that this is a problem that needs addressing. Critics of these initiatives argue that we should be asking whether data-driven systems shift power, not whether they are fair (eg Kalluri 2020). These critics are concerned that inequality and discrimination may be conceived as technical problems within FATE/FAccT debates, rather than the socio-technical problems that they are. And yet the two groups – critics and FATE/FAccT communities – are not mutually exclusive, with many people within FATE/FAccT circles taking a critical approach and recognising structural inequalities as socio-technical problems.

On *Living With* Data, we explored whether fairness matters to non-experts, and what fair data uses look like to them, in part because of these debates about the value of focusing on fairness. At the end of each

focus group and interview, we asked participants if they thought that the data uses we had spoken about were fair or unfair. If participants struggled to answer this question, we used alternative terms such as 'do you feel OK/not OK with them' and 'are you comfortable/uncomfortable with them'. We asked how important it was to participants that data is used in fair ways, to explore whether fairness matters in people's perceptions of data uses in different domains of everyday life.

4.1.1. Fairness means different (combinations of) things to different people

As Kennedy and others (2015) found, **the motivation** for gathering and using data and **the purposes** for which data is gathered were important factors in considering whether data uses were fair. This was raised in Mbali's interview, which focused on the theme of Algorithmic Processing. Mbali is a black, British-African, heterosexual woman, aged 25-34, a student with an annual household income £30,000-£39,000 and no long-term conditions. After a conversation about the various aspects of DWP Dynamic Trust Hub, BBC Box and BBC Own It, in response to the question 'what do you feel makes a fair use of data more generally?, Mbali stated:

> *I think it's the purpose of the use. We live in capitalist societies, so somebody is always after your money in some way or form [laughs], and in that way, I think commercially it's very difficult for your data to be used fairly, 'cos it is always in some way manipulative. But in terms of the public sector, I think, if they are not led by profit and they're led by trying to provide certain services, I think that changes it. […] I guess it's just about what the motivation is behind them doing certain things, and if that's fair, then the whole process should be fair.* (Mbali)

Tanya, a white, lesbian woman who was born in the UK, aged 55-64, who works in the public sector, has a long-term condition and an annual household income of £30,000-£39,000, made a similar point. In an interview about the Data Matching theme (which included the NHS antibiotic prescribing project and Confirm Your Identity), Tanya responded to a question about what makes data uses fair by describing them as:

> '*for the greater good […] for health and benefits, for people, for us to learn, develop and grow in a way that's going to make society a fairer place and people have the best opportunities in their life'.* (Tanya)

In this quote, Tanya describes data uses **for the public or social good** as fair. This is also implied by Mbali, when she says that data uses in the public sector, which provide services and are not profit-driven, could be considered as fair.

In addition, data uses which are not equally **accessible to all** were seen as unfair by Tanya. In a later conversation, she described the NHS antibiotic prescribing research project as fair because of her *'sense of where it's motivated from.'* In contrast, she felt that DWP Confirm Your Identity was not *'completely fair and all-inclusive for all communities.'* Tanya explained:

> *People don't have access to technology in the first place. And you know, what about people with English as an additional language or no English, or families with additional needs, where using technology might be difficult, or they aren't literate.* (Tanya)

Tanya made this comment after we had discussed the potential benefits and harms of Confirm Your Identity to her. After describing the Confirm Your Identity process and its benefits, we pointed out that some groups were concerned that identity checks are not accessible to all, such as people without credit

histories, with unusual residences or residence histories, or otherwise complicated lives. This might mean that such people are put under more scrutiny or have their access to welfare benefits delayed, something that is also a problem with non-digital, face-to-face or paper-based identity verification systems. We reflect on the relationship between our decision to describe such concerns and participants thoughts and feelings about data uses in Section 7 below.

What was considered a fair data use varied for participants. For some participants, an important component of fairness was **clear and honest information** which makes it possible to understand what happens to data. Lewis, for example, took part in an interview that focused on the Data Ownership and Control theme, covering the NHS Covid-19 Data Store, BBC Box and BBC Own It. Lewis is a white, heterosexual man who was born in the UK, aged 55-64, in work, with an annual household income of £50,000-£60,000 and no long-term conditions. Lewis said that the characteristics of fair data uses included:

> ***transparency and honesty*** *in terms of what's been captured, how long it's been kept for, and what purpose it's going to be used for.* (Lewis)

For Lewis, **being able to choose** *both* **to share data once you understand how it will be used** *and* **what happens to your data** were important components of fairness. He showed good understanding of data uses by mentioning the importance of knowing how long data would be kept for. Lewis carried on to say that once you have this information, then whether you can 'consent to it or not' or have 'the ability to opt out' are also components of fairness, drawing attention to the importance of **choice**.

Choice as a characteristic of fairness operates at different levels. Kahina, a Black, heterosexual woman who was born in Somalia, aged 25-34, a student with an annual household income £30,000-£39,000 and no long-term conditions, participated in a focus group discussion about algorithmic processing. She compared the two BBC examples with the DWP Dynamic Trust Hub in terms of choice, in order to evaluate their fairness. With BBC Box and BBC Own It, people *'have got a bit of a choice'* with regard to completing a profile or downloading them, she said. Because engagement with the DWP practice '*is a need'* for many people, and the potential security checks will not be optional, there is less choice, which, in Kahina's eyes, makes the practice more unfair.

Melissa, a white bisexual woman who was born in the USA, is 35-44, a psychotherapist with an annual household income of £30,000-£39,000 and no long-term conditions. She wanted to know whether choices could be made about uses of specific data, with regard to the NHS antibiotic prescribing research project. Talking about data about registered births and deaths, she said:

> *I'd want to know more about that, as to whether there was a choice involved in that, whether for the patients, you know, prior to them dying. Or if it was something that at least was, like, given to the family of, you know, we're trying to get this data, is it okay if we use this person and their circumstances to inform this. If there's a consent or choice involved with that, it would seem more fair to me. But if it's just kind of automatically used, because they're deceased, that makes me really uncomfortable and to me that's not very fair.* (Melissa)

Most participants recognised that **fairness is not just one thing – that is, not just clear and honest information, being given choices, being for the social good, or inclusive.** In most responses to questions about what makes data uses fair, participants discussed more than one factor. Being informed and understanding how data is used was often mentioned alongside being able to make a choice, or data being used for the social good. For example, in a focus group discussion about the Algorithmic Processing theme,

Louisa, a white, heterosexual female who was born in the UK, aged 35-44, who works for a charity, has an annual household income £40,000-£49,000 and no long-term conditions, compared the two BBC examples, BBC Box and BBC Own It, with the DWP Dynamic Trust Hub and talked about the equal importance of consent, choice and understanding what is being consented to:

> *I think if you voluntarily choose to share your data then that is a fair use of data. If your data is taken without your knowledge or passed onto a third party or it's not transparently told to you how your data's been used, then that's an unfair use. Equally, we've established, if you've got to read seven pages to say you accept terms and conditions, is that actually a fair use?* (Louisa)

In this quote, Louisa indicates that fair data use is not only about being given a choice; it's about being able to make an *informed* choice based on accessible information. Lucas made a similar point by highlighting what he considered to be unfair. Lucas is a heterosexual man from a mixed ethnic background who was born in Mexico. He is aged 35-44, with an annual household income of £10,0000-£19,000 and no long-term conditions. In a conversation about data matching, which included discussion of the NHS Antibiotic prescribing research project and DWP Confirm Your Identity, Lucas said that it's unfair when companies exploit people's lack of understanding of data '*collecting, sharing and analysing'*. He stated:

> *I think that what is unfair there is that these kind of companies take advantage of our ignorance, and that they profit from it'.* (Lucas)

Ahmed is a Pakistani heterosexual man who was born in the UK, aged 25-34, unemployed and looking for work, with an annual income of £9,999 or less and no long-term conditions. After a conversation about various aspects of DWP Confirm Your Identity in an interview about the Data Matching theme, he said that fairness involves clear information, transparency, data being used for the social good and the ability to consent:

> *It's where people are kept well informed and data is used for a valuable or useful purpose, and there's a transparent process as well for sharing the data. And obviously permissions are sought too.* (Ahmed)

For Ahmed and other participants, being able to choose whether to share data is an important component of a fair data use, but this needs to be accompanied by understanding how data will be used and knowing that its use will be beneficial to society.

4.1.2. Different understandings of fairness can make talking about fair data uses difficult

Differing understandings of the term fairness made it hard for some participants to talk about what fair data uses look like. But most participants wanted data to be used in reasonable ways. Gaynor is a British Chinese woman who was born in Hong Kong. She is 25-34, a web designer with an annual household income of £70,000-£90,000, heterosexual and has no long-term conditions. She participated in an interview about the Algorithmic Processing theme, incorporating the two BBC data uses and DWP Dynamic Trust Hub. In response to a question about whether they were fair like Kahina mentioned above, Gaynor also compared the amount of choice that people have about using these systems,. She noted that if extra security checks were introduced into DWP identity verification, it would be to improve security, so although systems users may not be able to opt into or out of these checks, their purpose was *'for [users'] own good.'* These nuances made it difficult for her to decide whether she would classify this data use as fair.

There were other instances of this. For example, Nigel, a white, heterosexual man born in the UK, retired, aged 65+, with an annual household income of £30,000-£39,000 and long-term conditions, participated in a discussion of the Data Sharing and Re-use theme. He compared the two NHS examples and BBC Own It:

> *The first one, the antibiotic prescribing, I feel quite comfortable with that. I don't have any concerns about that, really. I think that's a reasonable, you know, use of the data. The second one, the Covid one, I mean, generally, I'm happy with it. I think it's important for that data to be analysed. I suppose I've just got some, I've just got some misgivings about the involvement of commercial companies, really, and I suppose what their access to the data is, which we don't know, do we? [Laughs] And the last one [BBC Own It], it's funny actually, I suppose on the face of it, it looks okay in a way, but I just have, I don't know, it just feels, it doesn't sort of sit right really. I don't know.* (Nigel)

In this example, we can see the importance of participants' feelings in their assessments of the fairness of data uses. Nigel could see that aspects of BBC Own It were fair, but something about it makes him feel uncomfortable, even if he can't rationalise or verbalise it.

A small number of participants asked us what we meant by fairness when we asked them if they thought that the data uses we had discussed were fair. For example:

> *Fair in what way, how do you mean?* (Melissa, quoted above.)
>
> *It depends on how you define fair, doesn't it, really?* (Diane, who took part in an interview that focused on the Data Ownership and Control theme which included the BBC Box, BBC Own It and NHS Covid 19 Data Store. Diane is a white, heterosexual woman who was born in the UK. She is aged 55-64, with an annual household income of £40,000-£49,000 and no long-term conditions.)
>
> *I'm not sure what you mean by fairness in this context. I'm not quite sure what – how the word fairness fits into this. I'm not entirely sure what you mean. I mean, life's not fair, is it, you know. I mean, some people are born rich, some people are born disabled, some people are born poor, you know. I don't know what you mean by fairness in this context.* (Jim, European, a heterosexual man, aged 65+, with an annual household income of £40,000-£49,000 and no long-term conditions.)

From the quotes above, we can see that fairness is not always seen as a clear term, or one that is applicable to discussions of data uses, because there are many ways in which it is used. In these cases, the multiple, mutable meanings of fairness appear to make conversations about how it relates to data uses difficult.

However, after participants had questioned what fairness means or how it relates to data uses, subsequent conversations would often reveal how fairness was understood and that it did, in fact, matter to them. For example, Melissa, after asking *'fair in what way?'* talked about choice as a characteristic of fairness. After saying *'it depends on how you define fairness,'* Diane linked fair data uses to transparency. In these cases, even though our initial question about fairness stalled conversations, participants ultimately linked the term to things that mattered to them, such as transparency and choice.

In a focus group about the Data Ownership and Control theme (which covered the NHS Covid 19 Data Store, BBC Box and BBC Own It), Jim questioned *'how the fairness fits into this'* then started to talk about inequality as a form of unfairness, particularly in relation to income and disability. As seen in the quote above, Jim linked fairness to equality, but not to data uses. In the subsequent conversation, Jim stood firm on his position of *'not knowing what is meant by fairness in this context'.* However, as his friends discussed

their understandings of what constitutes a fair data use, Jim joined in, saying that he was concerned about *'less scrupulous'* people getting access to data. He went on to say that he was worried about the *'non scrupulous'*, which suggests that using data ethically or responsibly was important to him, even though he resisted the term 'fairness'. Despite these diverse and sometimes confusing meanings of fairness, we found that participants wanted data uses that avoided harms, even if they didn't seem the relevance of using the term fairness in this context.

4.1.3. There's a relationship between thinking about fairness and thinking about inequality

Some of our participants felt that data uses were fair if they did not affect different groups unequally. If data uses negatively affected or discriminated against already disadvantaged or minority groups, they were seen as unfair. As seen in the quote above, Tanya saw data uses which are not equally accessible to all as unfair. When asked whether they thought the data uses we discussed with them were fair, participants often considered how they might impact people from disadvantaged or minority groups, because of their income, limited access to the technical resources needed to access data-driven systems, or because English was an additional language. Without explicitly using the term, a lot of participants appeared to be aware that data uses can reinforce inequalities.

For example, Rosie, a white British heterosexual woman, aged 35-44, a physiotherapist with an annual household income of £30,000-£39,000 and no long-term conditions, participated in a focus group discussion about the Algorithmic Processing theme. When asked what she thought would make the data uses shown to her (the two BBC examples and DWP Dynamic Trust Hub) fairer, she said that information about data uses needed to be clear, concise and available in different languages in order to give people *'fair and equal access to understanding what they're sharing'*.

Kerry, a white, queer British woman who was born in the UK, aged 35-44, a public sector consultant with an annual household income of £70,000-£99,000 and no long-term conditions, took part in a different focus group about the Data Matching theme. For her, the welfare context of this data use was significant. Because already disadvantaged or vulnerable people are dependent on welfare for their survival, it is especially important that data uses in this context are accessible, she felt. Grace, a white, British, bisexual woman, aged 35-44, a teacher with an annual household income of £50,000-£69,000 and no long-term conditions, thought that the DWP Confirm Your Identity data use was unfair, because '*not everyone can access it*'. At the same time, she also recognised that other approaches to identity verification may also be inaccessible to certain groups:

> *Now that I've thought about certain people and certain, you know, groups of people that wouldn't be able to use that data matching system, there's definitely some unfairness with the DWP one. I just don't know if another system would be fairer or if it is a system that would suit everyone, really. […] I do understand that you can't just give Universal Credit to anyone with no proof of identity. It's maybe just inherently unfair.* (Grace)

Some participants linked their reflections about inequalities to fairness, voluntarily using the terms 'fair' and 'unfair' in their discussions without being prompted to do so by researchers. For example, Ruby, a heterosexual, British Chinese woman, aged 18-24, who works in legal services, has an annual household income of £40,000-49,000 and no long-term conditions, participated in a focus group about the Data Matching theme. Talking about the documentation needed to verify identity through the DWP Confirm Your Identity system, she said:

> *those people that don't have a P60, is it more people that have jobs that are cash in hand, for example, and people that don't have a passport, would they, you know, have never left the country or never needed a passport. That makes me think of people that may be [...] if you've never been abroad and have a cash in hand kind of job, that makes me think the people that are closer to the poverty line have to part with their data more than people that have like a P60, passport, like you said, and that's really unfair. (Ruby)*

Here, Ruby appears to believe that people without the right kinds of records to be able to verify their identity online would have to 'part with their data more' than people who have such records. Whilst this is not accurate, and we say more about misunderstandings below, Ruby accurately notes that already disadvantaged groups may be less able to engage with Confirm Your Identity than others.

In contrast, Tahira, a heterosexual, Pakistani woman, aged 45-54, a charity worker with an annual household income of £50,000-£69,000 and no long-term health conditions, thought that BBC Own It was fair because it could overcome inequalities in access to information and resources. In this way, she connected fairness with equality:

> *A lot of people I know wouldn't normally have access to that kind of resource. [Pakistani parents sitting at home in the UK] wouldn't know where to reach out to, because they've not been educated in this country, for example, or just don't know. [...] So, for me that's the fair one, if I was to look at it from that lens.* (Tahira)

These examples suggest that talking about whether data uses are fair or not may be how some people who do not have a political vocabulary – for example, of social justice, inequality, or power – express their concerns about the relationship between data uses and structural inequalities. In this way, the concept of fairness can be useful, as a way of ordinary people expressing their political concerns.

4.2. Inequalities and perceptions of data uses

Our second main research question on *Living With Data* was: *What do different people know and feel about specific data-related practices in different domains of everyday life?* We asked this question because we know that the context of data uses matters in people's perceptions of them, and so do other differences (we discuss these in our extensive literature review, which preceded our focus groups and interviews – see Kennedy et al 2020b). We also know that inequalities shape lives and lead to different experiences of data uses, with disadvantaged or minority groups more likely to be discriminated against in data-driven systems. A growing number of commentators are providing empirical detail about experiences of data-related inequalities, with the most prominent examples coming from the US, such as Virginia Eubanks' (2018) book *Automating Inequality*, Safiya Noble's (2018) *Algorithms of Oppression* and work by investigative journalists at *Propublica* (Angwin et al 2016) exposing how algorithmic systems in the criminal justice system discriminate on the grounds of race. Research into people's *perceptions of* data practices, as opposed to their experiences of them, engages much less with inequalities. This is starting to change, in part through our own work (eg Kennedy et al 2020a), but more understanding is needed of the role that structural inequalities play in people's perceptions of data uses.

On *Living With Data*, we found that:

- **Belonging to a disadvantaged or minority group** appeared to inform participants' perceptions of data uses and what they said about them;
- **Intersecting demographic characteristics** were important in participants' reflections on data uses.
- A number of **participants who didn't belong to a disadvantaged or minority group were still concerned about how these groups might be more negatively impacted** by data uses;

4.2.1. Belonging to a disadvantaged or minority group informs perceptions of data uses

We were interested to explore whether the demographic differences that result in discrimination inform perceptions of data uses. We therefore recruited diverse participants, as described in Section 3. We collected demographic data on each participant (summarised in Table 1), and analysed interview and focus group data in relation to this information. Belonging to a disadvantaged or minority group appeared to inform how participants reflected on data uses and what they said about them, in relation to: education; economic status; age; disability/long-term conditions; gender identity and sexuality; English as an additional language; race and ethnicity; and the intersections of one or more of these characteristics. To be clear, we are not suggesting that there is a direct correlation between belonging to a demographic group and attitudes to data uses. Our point is that demographic characteristics shape life experiences. In turn, those experiences shape perceptions of data uses. We provide some examples below.

Some of our **older participants** appeared to be more comfortable with NHS data uses than data uses in other sectors. Alan, a white, British, heterosexual man who was retired and had no long-term conditions, and his partner Betty, a white, British, heterosexual woman, also retired and also without long-term conditions, both aged 65+, participated in a discussion of the Data Matching theme, incorporating the NHS Antibiotic prescribing research project and DWP Confirm Your Identity. They were positive about data uses in the NHS Antibiotic prescribing research project. For example, Alan said:

> *If we'd been asked to cooperate in our hospital environment with a thing like this, we would have jumped at it, wouldn't we*? (Alan)

Betty agreed, and other older participants also concurred. For example, Bruce is a white, British heterosexual man, aged 65+, who is retired and has an annual household income of £20,000-£29,000. He said of the NHS Covid-19 Data Store (in a discussion of the Data Ownership and Control theme) *'I would have thought that the more medical information they have, the better they can target treatment'*.

Some participants with **long-term conditions** also felt the same. Indeed, Bruce himself has long-term conditions. Elsewhere, in a face-to-face focus group on the Data Matching theme with four women with long-term conditions, the NHS antibiotic prescribing research project and DWP Confirm Your Identity were discussed. These participants agreed that they trusted the NHS and therefore did not mind it using their data. Celeste, a black, British Caribbean, heterosexual woman born in the UK, aged 35-44, unemployed and looking for work, explained:

> *if something does happen, they should be able to type your name into the system and be like, 'Oh, this is such and such. She has recently come out of hospital for such and such', when, what she was prescribed, it should all be updated and in a place, I don't know where, but stored somewhere.* (Celeste)

In contrast, NHS and health data uses seemed to concern **LGBTQ+ participants** more than heterosexual cisgendered participants. A discussion of the NHS antibiotic prescribing research project took place between Todd, a white, gay man born in the UK, aged 25-34, with an annual household income of £9,999

or less and no long-term conditions, and his boyfriend Matthew, a white, gay man born in Zambia, aged 35-44, a web developer with an annual household income of £20,000-29,000 and no long-term conditions. Todd and Matthew reflected on the importance of the specific type of health data in question. They said that they would feel differently if the research project used data about *sexual* health, compared to data about colds or the flu, because of the ways in which this kind of health data can be mobilised for harmful ends. Drawing on his awareness of how data uses might enable discrimination structures, Todd said:

> *The sexual angle is a really, really good point, particularly when it comes to LGBTQ and that community, because we know that some of the information about our sexual health in the past has been used to massively discriminate, to massively sort of chastise those communities. It's interesting from that perspective that it could potentially take one person with a vendetta of some sort – or just a mistake, just a bit of human error, that could then cause ramifications for that particular group or population that maybe more – that therefore the risk might be heightened.* (Todd)

Ellis participated in an interview about the Algorithmic Processing theme. They are white, British, non-binary and bisexual, aged 25-34, a data analyst with an annual household income of £30,000-39,000 and no long-term conditions. Ellis was also cautious about health data uses. Towards the beginning of the interview, Ellis was asked whether they had ever changed their behaviour because of their knowledge or experience of data collection practices. Ellis said they had altered some of their social media activity to reduce producing data about their identity as a trans person, and they were selective about who they shared their health data with:

> *There's some things, especially NHS things, where I don't disclose honestly, like, medication, because, so, they'll ask you like are you taking any regular medication […] but if I perceive that the fact I'm taking hormones isn't relevant to them I won't let them know. So I'll just lie about it, even though it's probably confidential and you expect the staff to have training, I'll just be like, right, it's easier for me if I don't mention that and just try and not flag up being trans at all. Like at the opticians, or at the dentist, or like things like that. I get a bit paranoid that people will treat you differently and not necessarily in a bad way, just that they will realise and it will just feel, there will be a shift and they'll treat you different.* (Ellis)

Other LGBTQ+ participants expressed concern about the uses to which data about sexuality and gender identity could be put in other contexts. In a focus group conversation about data matching, which covered the DWP Confirm Your Identity and NHS antibiotic prescribing research project, queer couple Heidi and Kerry reported that when they had lived in a country in Africa where same-sex relationships are criminalised, they had been cautious about online activities that may have produced data about their sexuality. Kerry, mentioned above, said:

> *our approach to data and what we shared online and offline couldn't have been more different to how it is in the UK* […] *We didn't Google anything that would suggest anything other than that we were two interesting spinsters. It was like a really conscious thing. We had to be really careful. I think we just are maybe comfortable/complacent in the UK about data, and if the government changed and I found that we had some kind of hideous kind of, I don't know, right wing – more hideous right wing set up, I'd be much more careful about that.* (Kerry)

These examples from Todd, Ellis and Kerry show that the current or potential future contexts of data uses play an important role in people's perceptions of them and their subsequent actions. Context also

mattered to Tahira, mentioned above. In her case however, it was the broader context of **structural racism**, her anti-racist work and identity as a Pakistani woman that mattered. When talking about BBC Box in a discussion about the Algorithmic Processing theme, Tahira noted that everyday media consumption produces data which is fed back to platforms and organisations, which in turn recommend content based on what they perceive to be users' preferences. Giving the example of music listening, she said '*by listening to your regular stuff on Spotify you're telling them anyway [about your preferences]'.* She expressed concern about people being recommended media content that reinforces existing, racist ideologies through such mechanisms which can '*add to [...] the racism.'* Recommendations based on consumption habits means that audiences:

> *are not going to get like the other side [or other points of view] because this'll just constantly pump and drip feed them more. So I know that sounds kind of harsh. But unfortunately that's where my brain goes when it comes to these kind of algorithms about biases behind them.* (Tahira)

In this quote, Tahira linked particular kinds of data uses to the political environment, seeing them as actors in the reproduction of harmful norms, and she was not the only participant to do this. This example shows that Tahira understood that data plays a political role, reproducing inequalities, revealing a sophisticated awareness of the politics of data practices and data processes.

Participants from different demographic groups drew attention to the fact that understanding data uses is challenging (we say more about understanding in Section 6 below). These included **participants with long-term conditions, for whom English was an additional language, with low levels of education and other identity intersections.** For example, in a focus group about the Algorithmic Processing theme (which covered the two BBC examples and the DWP Dynamic Trust Hub), Richard and Lucinda, who both have **long-term conditions**, were asked whether they thought understanding what happened to their data was easy. Lucinda is a white, British, bisexual woman, aged 35-44, and is not in paid work. In answer to the question, Lucinda said *'oh god no, I don't know what happens'*. Richard, a white, British, heterosexual man, aged 45-54, and is dyslexic. He didn't think he had ever *'read through a single terms and conditions'* because doing so feels like *'an ordeal'* to him. He said:

> *All reading is just, not all reading, but just things where it's not designed to be clear, like in any way confusing [laughs], or legal jargon, or any kind of potentially ambiguous – I just [sighs], it does me in.* (Richard)

One participant, Brook, is a non-verbal communicator. He is a white, British, heterosexual male, aged 35-44, and in work. A facilitator translated his responses to our questions in his interview about the Data Sharing and Ownership theme. Towards the beginning of our interview with him, he was asked how he felt about data uses in society. He was generally positive, but was concerned about *'those who require the most assistance'* being *'adversely affected'*. Brook's facilitator highlighted the important role that his support staff play in Brook's life, enabling him to live independently. With regard to data uses:

> *Brook can opt out because we're here to have that conversation and to assist him in that. [...] [Other people who might not have such support] could end up being exploited in that. Because the onus is on them, so the person collecting the data will always say, 'But you didn't tell me no'.* (Brook)

For some participants for whom **English was an additional language**, understanding data uses was also challenging. Gulay is a Turkish-born heterosexual woman, aged 35-44, a student with an annual household income of £9,999 or less and long-term conditions. In a focus group discussion about the Algorithmic

Processing theme which included the BBC Box, BBC Own It and DWP Dynamic Trust Hub, she noted that she found understanding data uses hard, particularly as they are described in terms and conditions. At the same time, knowing that data about her was safe and secure was really important, because of her status as a refugee. She said:

> *This is hard for me because it's in English and long. I know I am not accept, not continue. I just accept. […] I'm refugee. I come to this country because of problems in my country. I don't want my location to be shared with this [Turkish] government. This is important for me. I know this country [UK] is very safe and – it is okay. [...] But I don't want to share with other countries, like my country [Turkey].* (Gulay)

Gulay is in the difficult position of needing to understand what happens to her data for her own safety, but struggling to do so because of the inaccessibility of explanations of data uses. Likewise, Lucas, a Mexican-born man whom we quoted above, acknowledged that having English as an additional language can make it hard to understand data uses. As a consequence, he said '*when you click on the accept all, [...] you have no idea what that all involves'.*

Like Gulay and Lucas, some participants who were or had been **on low incomes and had experience of applying for welfare benefits** noted that it could be difficult to understand or otherwise navigate data processes, especially in the context of welfare. Reflecting on DWP's Confirm Your Identity in a focus group about the Data Matching theme, Melissa, quoted above, combined her own prior experience of applying for welfare payments with an acknowledgement that a range of structural barriers may prohibit participation in online identity verification. She stated that having access to the internet, to a computer, a bank account and a home are all pre-requisites of being able to engage in online identity verification, whilst also noting that identity confirmation is important. Her own experience of not being able to provide all of the required documentation when she had applied for welfare some years earlier had made her feel *'almost like I wasn't good enough'* and she wondered *'how many other people are being excluded by that, or having that same feeling from it'*.

The **intersection of demographic characteristics** was important in many participants' reflections on inequalities and data uses. For example, above we saw that Gulay's status as a refugee intersected with the fact that English is an additional language for her, something which confounds the challenge of making sense of complex information about data uses. This led her to express concern about the safety of data about her and a desire to ensure it doesn't fall into the wrong hands. This example points to the importance of intersectionality, a term used by Kimberlé Crenshaw in 1989, to highlight how systems of power and inequality – for example, relating to gender, race and ethnicity, class – intersect (Crenshaw 1989). These intersections shape experiences of discrimination and privilege. For example, in a society that is sexist and racist, Black women experience discrimination because of both their race and their gender. For working-class Black women or Black women with disabilities, the cumulative effects of discrimination are even more acute. Such intersections inform how life is experienced, which in turn informs perceptions of data uses. Intersectionality is therefore a useful framework through which to make sense of perceptions of data uses.

Some older participants indicated that they struggled to understand data uses. Shazia, a Pakistani woman who is over 65, is not in paid work, has an undergraduate degree and no long-term conditions, required a lot of explanation in our interview with her about the Algorithmic Processing theme (including the two BBC examples and the DWP Dynamic Trust Hub). It wasn't clear that she had understood the purpose of or intentions behind BBC Box. Betty, mentioned above, answered the question 'have you ever become aware of how your data is used in everyday life' with a categorical 'no'. Her partner Alan, more confident

from the beginning of the conversation, wanting to guide her, referred to various online apps that they use on a daily basis, such as Spotify and Netflix. Ultimately, Betty had a lot to say about data uses in everyday life, including through her past involvement as a participant in NHS research projects and in a longitudinal study conducted by a University that tracked the health of women over time. She noted how she was '*very happy to help'* with this kind of work and thought it was 'very important' to collect this kind of data for research purposes. Shazia and Betty's lack of confidence in their ability to understand and talk about data uses could be explained by the intersection of age and gender – and, in Shazia's case, possibly having English as an additional language. The fact that Betty went on to open up about the topic highlights an important point that one of us, Helen, has identified in previous research with some of these demographic groups: that it is not necessary for people to fully understand data uses in order to have opinions about them (Kennedy et al 2020a). We elaborate on the importance of focusing on understanding in Section 5 below.

4.2.2. Participants who didn't belong to disadvantaged or minority groups were still concerned about how these groups might be negatively impacted by data uses

Some participants who did not belong to a disadvantaged or minority group were still concerned about how these groups might be more negatively impacted by data uses. Often, participants from one disadvantaged or minority group were concerned about the effects of data uses on another disadvantaged or minority group. These participants were aware of the relationship between data uses and inequalities, and that some data uses are more likely than others to deepen inequalities. They were most concerned about whether and how people from disadvantaged or minority groups might engage with the two DWP data uses, Confirm Your Identity and Dynamic Trust Hub. Accessibility was seen as particularly important with regard to these two examples, because DWP provides essential welfare services to the very people who are likely to be disadvantaged by structural inequalities. This confirms a central finding from the *Living With Data* survey, that context is a defining factor when it comes to public attitudes to data uses (Kennedy et al 2021). In other words, data uses in some contexts are more concerning than others, and people's attitudes to data uses differ, depending on the context. A small number of participants raised concerns about inequality in discussions about BBC and NHS data practices. These concerns tended to relate to access and accessibility or the circulation of data beyond its original context, rather than the specific data uses themselves.

For example, as already noted, in a focus group discussion about DWP Confirm Your Identity, Ruby, mentioned above, expressed concern about people who don't have documentation such as a passport or a P60, and worried that people living '*closer to the poverty line'*, who may have cash-in-hand jobs, might find online identity verification more challenging than she would. In a separate discussion of the Algorithmic Processing theme, Tahira, who is also quoted above, reflected on the extra security checks that were being considered as part of DWP Dynamic Trust Hub, which included checking time of log in, rhythm with which people type their passwords, swiping patterns and the devices that are used to log in. Although not on a low household income herself, Tahira's networks of people struggling with multiple jobs led her to be concerned about how difficult it would be to meet the requirements of these checks. She said:

> *Five percent of me says, yeah, I get why there is a need for extra security. But the 95 percent of me thinks it's bureaucracy and it's actually making the process more bureaucratic and more exclusive. People struggle anyway with online stuff. And also, you know, if someone's working three jobs – I'm just using that as an example, I know so many people that do that now, unfortunately, men and women. And so, yeah, they might do it in between shifts or might be on rotas or whatever. So yeah, to be managed or observed, because that's what this is, on the time of day that you might log on or*

the location. And then if it doesn't always match up, you're going to have to go through another layer of security. I know that's one example, but I could give you so many about how this would be so stressful for people. (Tahira)

Tahira also expressed concern about the effects that socio-economic inequalities have on parents' ability to engage with their children about their phone and app usage, in the context of a discussion about BBC Own It. Whilst she was positive about how Own It could provide access to resources to people who otherwise might not be aware of them, she was concerned that not all parents are equally able to support their children's mobile phone use, precisely because of these inequalities. She said:

One family member I know, like family friend kind of thing, she's on it. She will check the kids' phone, she will – you know, she is like really – without being too aggressive, she knows how to manage that. Whereas other parents are working three jobs, they haven't got the time or the, you know, know-how of what to check. Even if they were to ask that child, 'Oh, what are you doing?' They'd say, 'Oh, yeah, I'm just doing this'. And they're like, 'Alright'. (Tahira)

Some participants recognised that their own privilege gave them an advantage in engaging with data processes which other people didn't have. For example, in a discussion about the Algorithmic Processing theme, Rosie, mentioned above, described how complicated she and her husband had found Government Gateway, the online system for accessing government services which it is necessary to log into as part of Confirm Your Identity, when they had recently used it to apply for Universal Credit:

It took us forever and we're like two, we've both got degrees and have been through university and English is our first language. I dread to think for somebody who isn't, you know, who hasn't got English as a first language. It'd just be horrendous. (Rosie)

Rosie differentiated herself and her husband from other potential claimants, referencing their high educational qualifications ('*we've both got degrees*') and language ('*English is our first language*'). She recognised her own and her partner's privilege and showed empathy for others who are disadvantaged by structural inequality. Rosie understood that the additional Dynamic Trust Hub checks would enhance security but was concerned that these might make an already complicated process more difficult for particular groups.

Also talking about Government Gateway in the context of DWP's Confirm Your Identity and the Data Matching theme, Kerry, who we mentioned earlier, felt that the difficulties she encountered trying to use it would be exaggerated for single parents or people dependent on welfare benefits. Talking about herself and her partner, she said:

between us, we've got about seven degrees, but we spent one evening tearing our hair out, trying to navigate the Government Gateway. If I was a single mum, stressed out of my head trying to get money to feed my kids and I was faced with that, trying to verify myself on the Post Office app and then remember what you did last Tuesday. (Kerry)

In a separate focus group conversation about the Algorithmic Processing theme, three friends Huso, Hajeera and Kahina, compared DWP Dynamic Trust Hub with the two BBC data uses. Huso, a black, British African, heterosexual woman, aged 25-34, a teacher with no long-term conditions, acknowledged that making sense of complex information about data uses, which she described as '*just*

legalese' could be difficult for groups who, unlike her and her friends, did not have a university education:

> *if you're younger, maybe if you haven't been educated up to university level, if you don't speak English, if you have a learning difficulty, it makes it hard for you to understand.* (Huso)

She noted that people with limited financial resources who lived in her community accessed the internet at the library, which she felt would make undertaking Dynamic Trust Hub security checks difficult:

> *Who wants to be filling out security checks in a public library, where they can see what you're doing on the computer, because the computers are placed so that everyone can see what you're doing on the computer? You don't want everyone seeing all that information about you. But if that's the only place you have to actually use a computer, […] that's really bad* (Huso).

Like Huso, other participants recognised that multiple sources of disadvantage could result in data-related vulnerabilities. They also noted that the intersection of more than one disadvantage could make encounters with data processes especially difficult. For example Louisa, a white British woman who works part-time for a homeless charity, who we mentioned earlier, acknowledged the challenge of making sense of already complex data uses for people for whom English is an additional language, who may be refugees, and whose lives may already be complex because they are homeless and in need of benefits. In a focus group discussion, she mentioned all of these factors when reflecting on the DWP Dynamic Trust Hub:

> *I think the problem I could see with this from a work point of view is the people that I work with can be very chaotic, English can be a second language, so understanding the instructions – and then if they have to do extra checks and – could it disproportionately affect a certain type of people, who already find it hard enough to understand the whole DWP process, because it's horrible anyway, especially if you're a refugee and have to prove your identity in about fifty million different ways.* (Louisa)

Other participants also recognised the importance of identity intersections. For example, in a focus group of the Data Matching theme, which included the DWP Confirm Your Identity and NHS antibiotic prescribing research project, Tanya**,** whom we quoted above, imagined how inputting personal data into online systems might be especially difficult for people who were anxious about authority and who had criminal pasts which they had put behind them. She identified this combination of characteristics as being likely to make using data-driven systems challenging. Heidi, a queer, white British woman born in the UK, aged 25-34, works as a doctor, has an annual household income of £70,000-£99,000 and no long-term conditions. She reflected on the data-related vulnerabilities that emerge at the intersection of LGBTQ+ identity and other factors, including age. Heidi said:

> *we're also very fortunate that we're, you know, comfortable in terms of financially and socially comfortably, you know, that we haven't got a fear of how our society views, whereas I imagine still within the UK there is definitely a subset of people where, actually, their community wouldn't accept them being gay and actually that data and that information is then really precious. So actually if they are then Googling, you know, gay websites or something and then that flips into their advertising and their family member picks up their phone, actually that is then really important. And, you know, working with children as well, often it is early teen years when children*

start exploring their sexuality, their gender and all that thing and they don't want their parents to know, but actually that data is collected and then regurgitated back at them. So I think that's something that can be potentially really harmful because those parents can react in a bad way. You can get abuse, children that are kicked out from their homes, from data that's shared back. (Heidi)

In these latter examples (from Rosie, Kerry, Huso, Louisa and Heidi), participants differentiated themselves from others for whom they felt that understanding and engaging with data processes might be more challenging. They understood these differences to emerge from structural inequalities, relating to education, income, English as an additional language, age, and refugee status. All of the examples in this section show that there is widespread concern about the societal consequences of data uses, not just for participants themselves, but for people in disadvantaged and minority groups who may be negatively impacted by or excluded from data processes.

5. How people understand data uses

In many of the quotes in the previous section, it is evident that how data about us is collected, analysed, shared and used can be difficult to understand. Data uses are complex, explanations can be long and hard to follow, often by design, and sometimes they are not available at all. Making sense of data uses can be harder for some groups, because of their age, educational level, income, English as an additional language, or long-term conditions, as discussed above. Greater transparency or improved data literacy are sometimes proposed as solutions to people's struggles to understand data uses. Underpinning both of these suggestions is an assumption that more and better information and skills will improve understanding. But how do people come to understand data uses? Understanding is often an important prerequisite to developing opinions about things like data uses (although this is not always the case, as Kennedy et al 2020a note). Yet understanding is rarely the focus of discussion, in policy, practice and research circles. One exception is the work of one of us, Susan, whose book *Understanding Well-being Data* (Oman, 2021) and other work has unpacked the complexities of understanding data uses. On *Living With Data*, we didn't set out to research processes of understanding, but we discovered that it was relevant to our aims. In this section, we focus on how people understand data uses, highlighting the role that imagining plays in such processes, and the importance of everyday life experiences as routes to understanding.

We have already seen many of our participants acknowledge that it is hard to make sense of information which ostensibly aims to explain data uses. Gulay noted that she found understanding documentation about data uses such as terms and conditions difficult. In an interview about Data Matching, that focused on the DWP Confirm Your Identity and NHS Antibiotic prescribing research project, we asked Grace, a white, British, bisexual woman, aged 35-44, a teacher with an annual household income of £50,000-£69,000 and no long-term conditions, whether she thought it was easy to understand what happened to her data. She replied *'No, no not all'* and went on to explain:

I think it's some of the language that they use. If you're not involved in using data things or if you're not familiar, with it, I think some of the language is quite confusing. I think sometimes there's quite a lot to read. If you're trying to get onto a website, if you were to actually sit and read every bit of it and think about what all of that meant, then you would be there all day. And sometimes I do feel as though that's kind of the point as well, to put everything hidden and so that you click it. (Grace)

Indeed, given the complexity of data processes, a small number of participants felt that a lack of clear information meant that a data use was unfair, as discussed in the previous section. For example, in a focus group discussion, Lucas, mentioned above, said:

> *I think is unfair, because we cannot expect that we all know all those different layers involved in collecting, sharing and analysing data.* (Lucas)

Being able to understand data uses was seen by some participants as a characteristic of fairness, also discussed above. A minority of participants believed that finding data uses too complicated leads people to give up trying to understand them. For example, Ella, a white British, heterosexual woman, aged 35-44, with an annual household income of £9,999 or less and who has a long-term condition, said:

> *You're never going to fully understand or know where or what your data is being used for. I suppose in a way, a lot of us are quite blind and we don't want to know, we just take it as a normal thing, a not-really-thought-about kind of everyday thing. If you really do sit there and think about it, you could overthink it on every little aspect of it. It's like why do they want that little bit of information, or why do they want to know this, or I want to know why they want to know this, you know, you could go on for hours, days, weeks. You could drive yourself crazy trying to work it out and understand it. [...] I think you just put your shutters down.* (Ella)

And of course, sometimes people don't understand data uses. During our research, there were moments when our participants appeared not to understand the data uses we discussed with them, despite our efforts to produce simple accounts and clear visualisations. In our discussion of the relationship between fairness and inequality above, we saw Ruby believe that people without the right kinds of records to be able to verify their identity online would have to 'part with their data more' than people who have such records. She went on to say '*I'm still kind of hung up on the whole financial agency getting your informatio n'*. This suggests that she thought that data would be transferred from one organisation to another as part of this process, which is not the case. Rather, data held by one organisation is checked, anonymously, against data held in another organisation, to verify identity. Similarly Diane, a white British, heterosexual woman, who we mention above, said of the NHS Covid-19 Data Store:

> *Why did they feel the need to contact Amazon, Microsoft? What's the benefit? This is supposed to be NHS, so it's supposed to be, in my mind, medical information. Why does Amazon need to know that? Why?*

Amazon Web Services and Microsoft provided technical infrastructure for the NHS Covid-19 Data Store. At the time of the discussion, it was not known whether they were able to access the data as Diane assumes they could.

These examples suggest that the involvement of third party or commercial organisations in public sector data processes was particularly confusing for participants. In the case of the NHS Covid-19 Data Store, lack of clarity about who had access to data and the specific contractual arrangements with the commercial organisations which were involved was the main concern that we discussed with participants, so it is not surprising that they were confused about this. **Lack of clarity leads to confusion, which in turn leads to concern.** In our survey, we also found widespread concern about commercial involvement in public sector data infrastructure and about commercial organisations profiting from personal data which was originally gathered for public or pro-social purposes (Kennedy et al 2021). Further research is needed into the

specific aspects of commercial company involvement in public sector data systems that are concerning and confusing, the exact origins of confusion and concern, and what public sector organisations can do to overcome them.

Most of our participants didn't 'put the shutters down', Ella's suggested response to the effort required to understand data uses. Rather, they worked to understand data practices. We identify two main ways in which they do this: 1) by **imagining** what happens to data, how processes work and their impacts on others; and 2) by drawing on their **everyday experiences** to make sense of data uses. We discuss these two themes below.

5.1. The role of imagining in understandings of data uses

We found that imagining played an important role in participants' understandings of data uses. By imagining, we mean building or creating a mental image of something that is not present at the moment of expressing an opinion. We are not suggesting that the things that participants imagined were imaginary – that is, existing only in the imagination. Rather, participants imagined, supposed and assumed things to fill in assumed gaps in explanations. The fact that participants often imagined that there was more to data uses than what they saw and were told suggests that the link between information and understanding is more complex than is sometimes assumed. Clear and transparent information about data uses does not always result directly in understanding. We think that historical data misuses, security breaches and lack of transparency may lead participants to assume that there must be information missing from accounts of data uses, and they imagined to fill in real or imagined gaps.

Participants therefore sometimes looked beyond what we told and showed them in our accounts and visualisations of specific data uses, to imagine:

1. **What might happen to data in the future**, for example if governance arrangements or prevailing norms change. Data policymakers and stakeholders often advocate for transparency about data uses and data practices as they operate ***now***, but people are concerned about and imagine possible future uses of gathered data, knowing that data uses change over time.
2. **Differences between how data uses are said to work in theory, and how they actually work in practice.** Participants sometimes looked beyond what they are shown or told about data uses, imagining a gap between how data uses are said to work 'on paper' and how they work in real life, sometimes drawing on their own everyday life experiences.
3. The **experiences and perceptions of people from disadvantaged or minority groups to which they don't belong**, who may be more negatively impacted by data uses than participants themselves would be.

This last point confirms what one of us, Susan, has identified in previous research about people's understandings of data uses (Oman 2019). That is, in trying to understand how they feel about specific data uses, people imagine how they work or don't work for others, they put themselves in the place of others, considering whether data uses impact negatively on them. In Section 4.2.2., we discussed the third form of imagining. In this section, therefore, we focus on (1) imagining what could happen to data in the future and (2) imagined differences between the theory and practice of data uses.

5.1.1. Imagining what might happen to data in the future

Some participants were concerned that data **could be used in ways and for reasons not foreseen, intended or described** at the time that data uses are planned or implemented. Louisa, whom we quoted above, said the following about the BBC Own It app:

> *What if the research done though, produced by the university, by lovely people like [our LWD researcher], was used by people for the very opposite of what it was originally designed for, to target specific children?* (Louisa)

A number of participants expressed similar concerns about possible future uses of data which may result from a change of government or policy. Governments and political conditions can change, a number of participants noted, and this might lead to changes in who has access to data, how data is interpreted, and how it is used. Craig, a white, British, heterosexual man, aged 35-44, a web developer with an annual household income of £30,000-£39,000 and no long-term conditions, took part in a focus group that focused on Data Ownership and Control. Craig was concerned that as a result of as-yet-unforeseen data sharing, personal health data from the NHS Covid-19 Data Store could end up 'in the wrong hands'. The data could then inform decisions that impact negatively on groups of people. Craig used insurance as an example:

> *some data [could] get used to discriminate for insurance against one particular health condition, or pre-existing health condition – and I think that's very feasible even in England. If you're trying to get insurance and you fill in your questionnaire and […] they've done a mass analysis on this data to find out, well, if you've had COVID and you're of whatever demographic, we're not going to give you insurance. I don't have any confidence in the government not to go and allow unscrupulous companies access to that data.* (Craig)

Ellis, mentioned above, was concerned that within the BBC Own It app, what is considered to be risky behaviour, or requires an alert or flagging, depends on interpretation, which in turn is informed by the broader political context. They were concerned that a child interested in climate change and searching for environmental campaigning organisation Greenpeace might be considered to be interested in terrorism, because at the time of the discussion, Greenpeace was included on a counter-terrorism list, alongside neo-Nazi and other extremist groups. They compared this concern with a potential future scenario in which LGBTQ+ identity is challenged or denied, and thus searching for related content could also be flagged as risky or problematic:

> *I know the BBC are apolitical, I'd be concerned of what the BBC are going to be flagging up as a problem, and that depends on what society is. I guess what the BBC sees as a problem could change depending on political pressure and things. You get it in some countries where obviously like LGBT content is not allowed. That is probably a while off from happening in the UK, but if a child was looking at certain types of political stuff, would that just be flagged based on the fact it's not, you know, when all those organisations got put on the terrorism list, that were just like Greenpeace and stuff. I'd just be a bit concerned about who's making the decisions of what's okay and what's not.* (Ellis)

Todd's partner Matthew, mentioned above, had similar concerns. Talking about the UK Census, which collected data on sexual orientation and gender identity for the first time in 2021, he recognised that collecting such data was important to inform the provision of services. However, he was worried that as a gay man, he may want to protect data about his sexuality in a more dystopian future:

You know, if the government knows that I'm gay, for example, and a particularly right wing administration comes in, they know where I live, they know where I am. They can make it a legal requirement for companies to share that data with the government. They could very quickly arrest me or whatever. That has crossed my mind. I mean, it seems like a sort of dystopian fantasy but at the same time, it's possible. I mean, it's happened in the past in history and so it's the thing of you don't trust Facebook, but you do trust the government, but why? I mean, it's not as though politicians are morally superior to business people. They're often the same thing [laughs]. (Matthew)

In these quotes from Ellis and Matthew, we can see that being LGBTQ+ informed how they imagined the future, just as belonging to disadvantaged or minority groups informed how participants reflected on data uses more generally. Belonging to a disadvantaged or minority group also shaped imaginings.

Most of the time when participants imagined what might happen to data in the future, they were concerned about data uses under changed to governments, regimes or policies, with only a few exceptions. One was Mbali, a black British-African woman mentioned above. She speculated about how the **circumstances of users might change**, and the consequences that this might have. Mbali said of the security checks being considered as part of DWP Dynamic Trust Hub, which included comparing the time, location, device and rhythm with which a user logs onto a DWP system:

if your behaviour changes – let's say you get a stroke, or let's say you end up having to have somebody else help you, so they have to put in the information for you – the only thing I can foresee is that it might just lengthen the login procedure so that it becomes incredibly frustrating. (Mbali)

Very occasionally, participants imagined future scenarios where **more data sharing might be desirable**. Talking about BBC Own It in a focus group on the Data Ownership and Control theme, Diane imagined a scenario in which a child looked up how to make 'a bomb or something like that.' Diane wondered what would happen to this data, beyond an alert being sent to the child. In Diane's opinion, this action alone would not be adequate; she felt it might be necessary to alert more people than the child themselves.

5.1.2. Imagining a disconnect between how data uses work 'in theory' and 'in practice'

Participants sometimes imagined a disconnect between how data uses work 'in theory' and 'in practice'. They imagined there was more to the accounts of data uses than we presented to them. This helped them build a mental picture of the data uses, and it sometimes took them to a place of concern.

One reason that some participants felt that differences might exist between data uses in theory and in practice was because data-driven and algorithmic systems are **not guaranteed to be error-free**. Errors might lead to mistakes in how systems work, and mistakes have consequences. Talking about BBC Own It in a conversation about the Algorithmic Processing theme, Jill, a white, British, heterosexual woman, aged 45-54, who has an annual household income of £50,000-£69,000, was concerned that Own It might erroneously send alerts to children suggesting they had been involved in bullying, when they had not. She said:

Coming back to the thing we said about the algorithm and the fact that it's not human controlled and it could make mistakes. And you could have a child who's perfectly happy who suddenly gets an alert about bullying, and then freaks out and thinks that something's wrong and they've done something wrong. (Jill)

Algorithmic systems like BBC Own It might misinterpret user activity, Jill noted. She pointed out that certain words can take on alternative meanings when used by young people, citing the example of 'bad' being used to mean 'good'. Language is dynamic, and changes to language use after a system has been built might mean the system is not equipped to adapt to *'changing language, changing trends, changing behaviours'* as Jill put it. She continued *'I guess they would keep modifying it, but it's only ever as good as what it's been told to be'*.

In a focus group discussion about the Data Matching theme, Matthew, who has been quoted above, identified a couple of possible disconnects between how the NHS Antibiotic research project works in theory and how it might work in practice, both of which were based on personal experience. In the first case, he questioned the academic rigour which was said to underpin the NHS Antibiotic research project, based on personal experience of a disconnect between how data uses are said to work 'on paper' and how they work 'in reality'. This experience related to sexual health data, in which the fallibility of the human actors involved was the source of the disconnect. He had visited a sexual health clinic and had been asked for the details of previous sexual partners. He said:

> *I gave them details and then they said we won't be able to tell you whether or not they've contacted us, obviously, and I was like well yes, obviously. Then a week later, got a phone call, hey so and so hasn't called us. Did you give him your details? I was shocked because they were giving me personal details relating to somebody and, you know, they don't know whether this person, I mean, it could have been an abusive relationship, where that information could have had consequences for my safety, or their safety.* (Matthew)

He went on to link this experience to the NHS Antibiotic research project:

> *So, there is that. I mean, on paper this is great. On paper this is helping antibiotic research, you know, whatever. In reality, somebody who just has a target to meet for contact tracing doesn't care or doesn't know, considers me like, oh well he seemed nice. We can tell him, he won't mind. You know, I was nice. I'm not going to go spreading that information. But they don't know that.* (Matthew)

This real-life experience of a data confidentiality breach led Matthew to imagine that such incidents could happen again: his experience was unlikely to be an isolated incident. In the context of sexual health, he realises that the potential ramifications of such breaches could be harmful. '*Obviously STDs are seen very differently from the flu or a cold, that data can be used maliciously as well, and it's the safety around that data,'* he continued, drawing on knowledge of how sexual health data had been used to discriminate against LGBTQ+ people in the past. Drawing on everyday experiences to reflect on data uses, as Matthew does here, was common amongst our participants. We say more about this in the next section.

This experience led Matthew to question the reliability of other human actors in health data systems. Later, the fact that a friend works in a related health department made him reflect on whether all stakeholders would be equally committed to ensure that claims about data privacy and confidentiality would be upheld. He said:

> *I can imagine doctors and nurses are very passionately concerned about data privacy, whereas a receptionist might just find it funny, you know, or might not have that same kind of commitment. Then the other thing is it's like somebody you know. I mean, one of my friends works in phlebotomy, you know, testing samples. There is a possibility that […] something's labelled with my date of birth and she knows my date of birth. She would be able to go like, that's Matthew's date of birth, and make that*

connection. So it also relies on individuals and we all know individuals are not necessarily trustworthy or, you know, they're fallible. (Matthew)

Lewis, a white British man quoted above, had similar concerns about possible privacy breaches, this time in relation to BBC Own It. Taking part in an interview that focused on the Data Ownership and Control theme, he imagined a disconnect between the aim of the data practice (to protect children's mental health) which *'on the face of it sounds great'* and how secure it actually was. He said:

> *You know, you could imagine if there was a way for somebody to hijack it then it would be an absolute nightmare in terms of, you know, it's now capturing all of the keystrokes from somebody, and in this case it's a kid so it's unlikely to be, you know, military secrets or whatever, but it's still that kid's kind of personal life that they are capturing in some way. So yeah, so I think – so this one on the face of it sounds great, but you'd really want to know a little bit – well, quite a bit more actually about that.* (Lewis)

5.2. The role of everyday experiences in understandings of data uses

In some of the examples in the last section, participants turned to their everyday life experiences to 'fill the gaps' between the accounts of data uses that we shared with them and what they imagined happens in practice or might happen to data in the future. Sometimes these imaginings were informed by participants' identities and the discrimination that they or people like them had experienced. All of them show that everyday experiences are an important resource that participants drew on in order to understand the specific public sector data uses that we discussed with them. Participants linked everyday moments to wider social factors and structures, and their observations about the role of data in their everyday lives sometimes served as opportunities for reflection on the politics of data uses or for small-scale acts of agency. We saw this in participants' stories of Covid-19 track and trace and vaccine trials, personalised advertising on social media, visiting the GP and data uses in a boxing class, to give just a few examples.

At the start of our focus groups and interviews, we asked participants about their awareness of data uses in everyday life, as a way of getting our conversations about specific data uses started. Answers to this question showed that many participants had wide-ranging awareness of data uses in a variety of everyday settings and situations. In stories about everyday encounters with data, some participants demonstrated their agency to everyday data scenarios. We provide examples of this below.

In a focus group discussion of Data Sharing and Re-use, Shani, a white, British woman, aged 25-34, who has a part-time administrative job and a long-term condition, understood that she received cold telephone calls from unknown organisations because her personal data had been shared with them. She didn't understand how this happened, and she responded in a number of different ways. Sometimes she didn't answer, but she felt she couldn't do this all of the time, because as someone who has learning difficulties and regularly attends hospital appointments, she needs to respond if there was a chance the call was from a doctor. Sometimes she provided the caller with false data, demonstrating her agency:

> *Sometimes it can come up as [number] withheld, which could be a doctor, so I feel like I have to answer it, and then you get all these other calls, like, "Hello, I hear you've been involved in a car accident." But to try and make those people waste their time, I try and play along with it, and mention all the Disney characters [laughs]. […] Just say, "Yeah, Donald Duck got whiplash," and, "Minnie and Micky got killed". […] When I'm not sure who it is, I won't answer. So yeah, it's basically just – I think we just need to know more from them as to what they're going to do.* (Shani)

Also responding to our general question about data uses in their everyday lives, Lucinda and Richard, a couple on a low income and both with health conditions, mentioned above, participated in a focus group on the Algorithmic Processing. They showed that they were aware of how data was produced through processes like registration at their children's kickboxing club. They also demonstrated agency in their response, refusing to participate in fingerprint registration, which they felt was unnecessary in the context of their children's small club:

> *Our kids go to kickboxing and it's a really small kickboxing place and they used to have this card system so you just, your name's on a card and tick it off like a register in school and all of a sudden they implemented – they must have been sold it by someone, you know, crazy salesman – but they suddenly implemented this fingerprint system. And the kids were really little, you know, they were like what twelve, ten, something like that, and I was like I don't want them fingerprinting, you know, I mean there's about fifty people or whatever that use this gym, why would they be using a system where they have to take people's fingerprints and store this information for some company so that they can sign in? When just signing, ticking them off in a box was perfectly adequate for the size as well of the organisation. Was just completely insane. So we wrote a letter refusing to do it and said, well we're really uncomfortable with this and we need an alternative, so our kids carried on just being ticked off.* (Lucinda)

This quote from Lucinda shows how the context in which data is gathered or used informs people's perceptions of them, something we we found in the *Living With Data* survey (Kennedy et al 2021), and which we also mention above in this report. In this instance, the fact that the data gathering process involves children, and Richard and Lucinda's belief that a fingerprint system was not necessary in a small club, led Lucinda and Richard to question and ultimately resist it.

In the examples above, participants responded to our question about whether they were aware of data processes in their everyday lives. Elsewhere, some of our participants talked about their everyday life experiences in response to questions we posed to them about the specific public sector data uses that were the focus of our research. In so doing, they linked *our* interests to *their* lives and experiences, to what matters to them. Because of this, their experiences of commercial companies' data processes sometimes informed their thinking about public sector data uses. Talking about their everyday experiences instead of the data uses that we presented to them could be seen as participants going off on tangents, talking about things that were not relevant. Instead, we suggest that linking our questions and examples to their everyday lives, participants talked about data uses in ways that 'connect data back to the social and political reality from which they were produced,' something that Catherine D'Ignazio and Rahul Bhargava argue is a necessary starting point in community-focused data visualisation learning (2019, p. 131). On *Living With Data*, we found that this is an important way that people come to understand data uses.

For example, Yasha, a heterosexual, Pakistani woman born in the UK, aged 35-44, a full-time mum with an annual household income of £9,999 or less and no long-term conditions, participated in a conversation about BBC Own It (as part of the Data Ownership and Control theme). She talked about her children's mobile phone use. Her oldest daughter, who was almost 13, had a phone so she could let her mum know when she was safely on her way home from school. Reflecting on her daughter's phone use and bringing detail about her daughter's everyday rituals of travelling to and from school into the conversation helped Yasha to make sense of BBC Own It. It helped her come to the conclusion that Own It was more appropriate for younger children, who might be less mature than a 13-year-old in their mobile phone and app usage. When the researcher tried to bring the conversation back to the specificities of how Own It

works, Yasha drew on another, different life experience. She explained that as a previous victim of fraud, she was cautious about sharing personal data. She was willing to share data for '*something legit'* like NHS or government, but not for what she described as *'rubbishy things that I seem to be signed up for.'* Yasha drew on her life experience to evaluate Own It, and the fraud incident made her feel that data sharing in the context of Own It was not a life necessity.

Similarly, in a face-to-face focus group with four women on low incomes which focused on data matching, participants linked aspects of the NHS antibiotic prescribing research project to their wider lived experiences. An explanation of the terms 'anonymous' and 'de-identified' led participants to talk about data security breaches, relating to health data, child benefit data, shopping and banking. This in turn led to a discussion of NHS data hacking, as the NHS was the context for the antibiotic prescribing research project. One participant, Patricia, a white, British, heterosexual woman, aged 55-64, who has a long-term condition and an annual household income of £10,000-19,000, recounted that she had recently been informed about a sudden change in the availability of her personal health data to the dermatologist who she has been visiting since she was a small child. Patricia said:

> *I've been going under the dermatologist since I was two, but all of a sudden they haven't got all my records. So, what do they do with those then? Burn them?* (Patricia)

The group linked this example from Patricia's everyday life, which appears to be something of a tangent, back to a discussion of data sharing in the health sector, with some focus group participants concluding that Patricia's case was *'a very good argument for data being automatically shared, kept on records'.* Talking about everyday, tangential experiences of data enabled this group of participants to come to conclusions about what appropriate data uses might look like. These apparent tangents also tell us about participants' concerns and the kinds of things that people think about when they evaluate data uses. The tangents, including 'horror stories' of data misuse and theft, feed into, frame and inform how people understand the data uses we discussed with them.

Toby, a white, British gay man, aged 35-44, with an annual household income of £70,000-£90,000 and Paolo, a gay man born in Brazil, with a mixed ethnic background, aged 35-44, an annual household income of £40,000-£49,000 and no long-term health conditions, participated in a focus group discussion where they also discussed the NHS antibiotic prescribing research project. They reflected on the importance of knowing what data about them was being used for and who was in charge of making decisions, for example about what data to anonymise and what data to share. Diverting from the example of the NHS antibiotic prescribing research, Paolo linked the discussion to his own experience of filling in consent forms at appointments for a Covid vaccine trial in which he had participated:

> *I actually have a good example on when the data is changing, as a project goes on [...] Every time I go there for a follow up, there's always the same form. I need to fill in the form again because one of the disclaimers have changed, they've added one extra, you know, your data is going to be shared with someone else. I think I've signed at least five different disclaimer forms with added points throughout, in the last six months already. So they are constantly changing the specific bits of the research project, because they may have identified that that needs to – I'm just so bored of it [laughs], I just sit on the chair, like, what's changed this time.* (Paolo)

In this quote, Paolo's experience of changing data uses informed his view that how data is used can change, and he uses this example as a way of backing up his view that it's important to know how data about us is used. Attending Covid vaccine trial appointments, although ostensibly extraordinary, became

part of Paolo's everyday life for a short while, and the mundane experience of filling in ever-changing consent forms became a means through which he could make sense of other health data uses.

6. Perceptions of data uses can change

Perceptions of data uses are not static. They change for a number of reasons. In the sections above, we have seen that participants were aware that what happens to data can change, as in the examples from Paolo and Patricia, in which decisions about how and where to share data appeared to change. We also saw other participants acknowledge that governments, policies and prevailing norms can change, and this can mean that data uses in the future are different from data uses as described in the present, in theory and on paper.

For example, Kerry told us about how, living in a geographical context in which same-sex relationships were criminalised, she and her partner Heidi changed their view of what would constitute acceptable data-producing activities, and were cautious about online activities that may generate data about their sexuality. Several participants noted that their level of comfort with data processes in the here and now would change if the broader cultural and political context was to change. Matthew and Ellis both expressed concern about the possible future misuse of data about sexuality and gender identity, and Jill noted that if algorithms can't keep up with the ways in which language changes, they become simultaneously less effective and more concerning.

We discussed these examples above in order to highlight the various roles that imagining plays in the process of coming to understand data uses. They also show us that perceptions of data can change. As we have highlighted throughout this report, and in our report on our survey (Kennedy et al 2021), context plays a significant role in shaping what people think about a data use. If contexts change, perceptions may also change. This highlights the importance of ongoing research into public perceptions of data uses, as they alter in line with contextual change.

Forming opinions about data uses involves having some understanding of them, but it also involves **feeling.** Qualitative research into perceptions of data uses, including our own, has drawn attention to the important role that feelings play in shaping perceptions. For example, Helen, writing with Rosemary Hill, has argued that emotions play a role in how people make sense of data, and that data are 'as much felt as they are experienced cognitively and rationally' (Kennedy and Hill 2017). Emotions play a role in the formation of attitudes and the expression of perceptions, in relation to data uses as with other phenomena. Other focus group research by Kennedy and others (2020a) found that people do not need to fully understand data practices in order to have strong feelings about them. Our *Living With Data* research adds to this conversation by highlighting how **feelings can change over time**.

Diane, quoted above, is one participant whose feelings changed during the course of her focus group. Reflecting on the NHS Covid-19 Data Store shown as an example of the Data Sharing and Onwership theme, Diane initially felt positive about it, saying *'it looks quite impressive'* and that she *'can't see a problem with it anywhere'.* Diane's friend Tess was also in the focus group. Tess is a heterosexual, white, British woman who was born in the UK, aged 45-54, with an annual household income £70,000-99,000 and no long-term conditions. Tess expressed concerns about the potential for mistakes to be made given the quantity and diversity of data held in the Data Store. This led Diane to start questioning it for herself:

> *It's almost too big, isn't it? If they'd said, right, we're going to collect this information for that reason, then that's fairly simple. But what's happened, it's like a spider's web. And the trouble is, then you lose, you know, on these little spider's legs, you lose a) what it's all about and b) it opens up to, first of all it going wrong and not working properly, and people then getting the information for means that they shouldn't have. You know when you look at it in more detail [...] it's, you know, it does make you sort of change your mind.* (Diane)

Diane acknowledged that thinking about the details of a data use – in this case, the involvement of commercial organisations and the amount of data in the data store – led her to change her mind. Reflecting on uncertainties regarding the nature of commercial organisations' access to data in the store, which was not known at the time, Diane said *'it is a worry, to be honest'*. Within the short duration of this focus group discussion, Diane changed from feeling positive towards the NHS Covid-19 Data Store and not *'seeing any problems,'* to feeling worried about its uses of people's sensitive health data.

Billie is white, British, non-binary and queer, aged 18-24, with an annual household income of £10,000-£19,000 and no long-term conditions. Billie's feelings towards BBC Own It changed during the course of their interview about the Data Ownership and Control. Billie was initially concerned about it, having recently watched a programme about concerns relating to YouTube's algorithms, and because Own It was targeted at children. However, after considering the details of how Own It works, Billie said that they *'like the fact that the keyboard allows a lot of control'* because they *'very much like the idea of children having autonomy of themselves'*. They continued:

> *I feel like this could be very, very positive for children, especially if it's got things like wellbeing advice on how to manage your mental health and you know, reminding kids to take breaks from their phones and their tech, all technology, their consoles and anything like their laptops and stuff. You know, go outside recommendations and like that, trying to push kids to doing more a variety of stuff.* (Billie)

Having said this, Billie then wondered if there was a disparity between the theory and practice of what happens to data in Own It, something discussed above. Billie said that the *'execution'* of the BBC Own It application was what really mattered. The Own It algorithm, they said, needed to be '*watertight*', children need to be able to switch off the app when they wanted to, and any sharing of Own It data with university researchers needed to be transparent, including information about who was funding the research. Billie summarised their 'feeling journey' with Own It as follows:

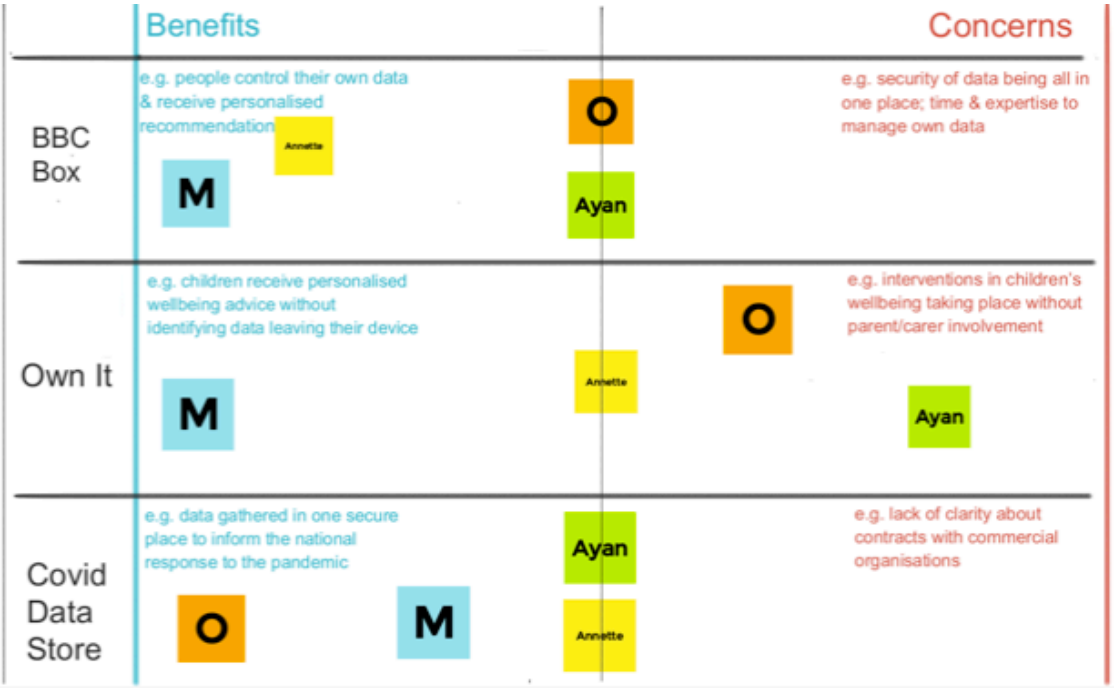
> *Initially I started off being very much like, oh god, children data harvesting, what is going on here, and then to being like that could actually be very, potentially very helpful, but it's like again, how is it doing this?*

Like Diane, in the short space of one interview, Billie's feelings towards Own It fluctuated between negative and positive. Billie's and Diane's changing feelings also show that it is possible to have contradictory feelings about data uses, which can be felt simultaneously, as seen in Billie, who is both positive and concerned about BBC Own It.

We also saw that participants' feelings about data uses changed as a result of the proposed benefits and potential harms that we introduced to them. As stated in Section 2, we informed participants about one claimed benefit and one claimed harm for each data use, because we believe that people can't assess potential benefits and harms if they don't know what they are. Participants were asked to consider

whether proposed benefits or potential harms of data uses mattered most to them. To do this, they placed Post-It Notes on an interactive whiteboard, an example of which can be seen in Figure 1 below.

Figure 1: A jamboard example with proposed benefits and concerns of data uses that we shared with participants



Below, Ellis, who we have mentioned above, is talking about where they placed their Post-It for BBC Box. The proposed benefits for BBC Box which were mentioned were that it gave individuals control of their own data and that, if they chose to share their data with media providers, they would receive personalised recommendations. The potential harm was that, as with all recommender systems, the system might recommend a narrow range of content. Commentators believe such systems may have troubling consequences for democracy, because they push people into echo chambers which reinforce their existing views and limit healthy debate. Ellis said:

> *I was going to say I'd put it [the post it] in the middle because you've still got control over it even if the control isn't detailed enough. But actually, I hadn't considered this concern and that's quite a good concern, because often on TV you see adverts that have clearly been tailored to you. For example, I watch a lot of Gogglebox, because it's great, and I always get, like, the car advert at the beginning, I always get the LGBT version of it, and I'm like, oh yeah, that's really good, but then I remember that actually probably I'm only being shown that because I fit the demographic. […] I'm thinking oh everyone's seeing this so it's good for representation, but actually not everyone is going to be showing that advert I don't think, and if you're only going to be recommended things that are in your sphere then that's not really very good, I don't think.* (Ellis)

Another participant whose feelings about a data use changed because of benefits and harms that we discussed with them was Grace. Grace is a white British, bisexual woman, who we mentioned above. Grace talked about where she placed her post it for DWP's Confirm Your Identity and why. As noted, after describing the Confirm Your Identity process and its benefits, we pointed out that some groups were concerned that identity checks exclude some people, such as those without credit histories, with unusual residences or residence histories, or otherwise complicated lives, and that this might mean that such

people are put under more scrutiny or have their access to welfare benefits delayed, something that is also a problem with non-digital, face-to-face or paper-based identity verification systems. Grace said:

> *I didn't really think of some of the concerns that are listed here, you know, excluding those without certain accounts. And, I suppose, you know, asylum seekers […], I don't know if they would have a bank account, so I don't think they would use that. I didn't really think of it until now. But that is a concern.* (Grace)

These examples show that feelings about data uses can change when new information is presented to us. Although the changes to feelings described here were a result of our methodological choice to talk about benefits and harms, in everyday life, we also become more informed about things like data uses by talking to other people, or coming across or seeking out information, and our feelings and opinions may then also change. In short, feelings towards data uses are not static. They can change in the space of a short discussion, they can change because contexts change, and they can change when people acquire new knowledge. This has implications for policy and practice, which need to be responsive to people's changing feelings about data uses. It is also important to acknowledge that research such as ours captures perceptions and opinions at a specific moment in time. Policymakers and practitioners building on the findings of research like *Living With Data* need to understand that perceptions and opinions are dynamic.

C. CONCLUSIONS

7. Reflections on talking to people about data uses

Some researchers claim that their participants do not mind that data about them is shared and used, whereas others find concern and criticism. On *Living With Data,* we asked how such contrasting findings might occur. Our evidence review (Kennedy et al 2020b) found that how research is conducted makes a difference to what it finds. Methodological choices, the questions asked, how findings are interpreted and presented, the disciplinary background and the political orientation of researchers played a role in shaping findings that emerge and claims that are made in the research we reviewed. In other words, how we talk to people about data uses matters, because **the methods we use influence what we find.** We recognise, like John Law and John Urry, that methods shape findings, because 'they have effects; they make differences; they enact realities' (Law and Urry 2004: 392-3). We are therefore aware that our choices have had effects on our findings, and we think this is also the case for those researchers who claim to find that people are not concerned about data uses. We talk about three important choices we made below, relating to research design, recruitment, and analysis.

Talking to people about their views on data uses is tricky, because data uses and processes are complex, abstract and not always of immediate concern. As described in Section 3, we made a choice to share accounts and visualisations of specific data uses with research participants, in order to make details about potentially unfamiliar data processes accessible to them. We did this because we believe that showing and telling people about data uses makes it possible first to understand them and then to develop and express opinions about them. We told participants about one proposed benefit and one potential harm of each data use, because we think that people can't assess potential benefits and harms if they don't know what they are. We think that it is methodologically unsound to conclude that people don't care about something without having taken steps to ensure that they know about it, although we recognise that understanding is complicated, as we discuss above.

Did we lead participants to express concern because we told them about the potential harms of data uses? It is possible that we sometimes did. On some occasions, our introduction of possible benefits and harms led participants to change their minds about a specific data use, as discussed in the previous section. It may be that this choice of ours meant that participants reflected on inequalities more than they would have done if we had approached our research differently. However, it should be noted that the majority of concerns and most discussions about inequalities occurred before we turned to the topic of benefits and harms, which was introduced towards the end of the interviews and focus groups. It is also important to note that only one of the six potential harms that we mentioned related to inequalities. Choosing to include statements about benefits and harms in our accounts of data uses, we sought to 'find the sweet spot' between informing our participants about data uses whilst not leading them towards particular conclusions.

Our decision to ensure that people from disadvantaged and minority groups were well-represented amongst our participants may also have influenced what participants talked about. As highlighted throughout this report, structural inequalities appeared important in multiple aspects of our research – in discussions of fairness, in participants' imaginings of future data uses or of disconnections between the theory and practice of data uses. We have highlighted how some participants who did not belong to a disadvantaged or minority group were still concerned about how these groups might be more negatively impacted by data uses. People who *did not belong* to a disadvantaged or minority group were in the

minority amongst our participants, given that we recruited with questions about inequality in mind. It was often the case that participants who belonged to one disadvantaged or minority group were concerned about the effects of data uses on people in other, different disadvantaged or minority groups, something we may not have found if we had recruited more participants from advantaged or majority groups.

We also made analytical choices which have influenced our findings. We chose to devote analytical attention to what might be seen as tangents, for example when participants answered our questions about specific public sector data uses by talking about their everyday life experiences. We see these reflections not as tangents but as important moments in which participants link what we want to talk to them about to their own lives. This enabled their meaning-making, as they transformed data uses from something complex and abstract to something embedded in their everyday lives. This choice, to listen to tangents, also has implications for what we found.

The particular accounts and visualisations and the specific data uses that formed the focus of our research also shaped our findings. As we noted in Section 2, we don't think it's possible to produce objective accounts and representations of data uses, because interpretation takes place as we describe or illustrate them. Instead of objectivity, we aimed for accuracy and balance. Our decision to work in this way has developed our thinking about what 'good' transparency about data uses looks like. Participants said that the visualisations we showed them helped them to understand the data uses we discussed with them, and they felt that organisations communicating about their own data uses could adopt a similar approach. They felt that organisations communicating to users about their own data uses could adopt a similar approach, and that this would be a more effective form of communication than current, lengthy Terms and Conditions. This simple step could significantly improve people's understanding of data uses.

Our extensive, iterative efforts to produce balanced and accurate accounts of data uses has led us to arrive at further conclusions about good transparency, which we think contribute to debate about this topic. We believe that good transparency communicates meaningful information about data uses to diverse publics, rather than being a PR exercise or designed to tick a transparency box. We think that good transparency should aim to contribute positively to enhancing people's collective agency and reducing power asymmetries in relation to data uses. We have developed a set of principles for good transparency, or what we call *critical transparency practice*, which we write about in a journal article.

8. Main findings and recommendations

It's important to remember that we are researching public perceptions of data uses in order to improve data uses, not to improve perceptions. Understanding public perceptions is necessary so we can then turn our attention to data uses and data-driven systems and think about how they should change. The recommendations below identify what data policymakers, practitioners and researchers can do to improve specific data uses and the data-driven ecosystem more broadly.

8.1. Address inequalities

There is widespread concern about the potential for data uses to have negative consequences for people from disadvantaged and minority groups and to reinforce inequalities, and people don't want that. This is are more concerning in some contexts than in others: for example, there is more concern about data uses reproducing inequalities in welfare than in other public sector contexts. Data policymakers and practitioners need to acknowledge differences, in data uses and in public views about them, and avoid simplistic understandings of public attitudes.

- **Recommendations:**

8.1.1. Avoid generalising about public opinions about data uses. There is no, one 'public,' and the idea of a singular public obscures the differences and inequalities that characterise 'diverse publics', which we suggest is a better term.

8.1.2. Understand the potentially discriminatory impacts of different data-driven systems, in order to a) communicate clearly about them and b) seek to overcome them.

8.1.3. When communicating about data uses, provide honest information about potential harms, especially for disadvantaged or minority groups, because this concerns people.

8.1.4. Regularly consult diverse publics, because data uses and perceptions of them change, and because structural inequalities influence what different people think of different data uses.

8.2. Ensure data uses have fair characteristics

Data uses in the public or social interest, that are supported by clear and honest information that is easy to understand were seen to be fairer than data uses without these characteristics. Fair data uses enable choice, about whether to share data and about what happens to it, and they don't negatively impact some groups more than others.

- **Recommendations:**

8.2.1. Consider not engaging in data uses that concern people, for example sharing data intended for pro-social or the public good with commercial companies who will make a profit from it.

8.2.2. Build choice into data processes, for example about what data is collected and what happens to collected data.

8.3. Be aware of and respond to imaginings

The role of imagining in people's perceptions of and attitudes towards data uses has important implications for policy. More information about data uses will not necessarily result in more understanding of them, but addressing what matters to people in explanations may help. *How* data uses are communicated to diverse publics is as important as ensuring they *are* communicated. Communication should be motivated by a genuine desire to enable understanding, and it should communicate information that is of interest and meaningful to diverse publics. Transparency efforts need to make it possible for people to understand what happens to data. They should speak to people's concerns about what might go wrong, what might change in the future and how disadvantaged communities might be affected. Historical data misuses, security breaches and lack of transparency may lead people to assume that there must be information missing from accounts of data uses, but empirical evidence of whether this is the case is needed.

- **Recommendations:**

8.3.1. Communicate clearly about data uses in order to actually communicate, not just to tick the transparency box. Explanations need to focus on what matters to people, for example: addressing what might go wrong (responding to the imagined gap between data uses as described on paper and in practice); possible future as well as actual current uses of data (responding to imagined future changes); and whether and how some groups might be more negatively affected by data uses than others (responding to participants' imagining the experiences of more disadvantaged others).

8.3.2. Commit to supporting a change of culture around data uses, in which data is not misused, security is not breached, transparency is sustained and purposeful, because what has happened in the past can influence how people perceive data uses in the present.

8.3.3. Use visuals to communicate data uses and ensure explanations can be easily translated to other languages. These simple steps could significantly improve people's understanding of data uses.

8.3.4. Support or commission further research to improve understanding of imaginings, where they come from and how to address them.

8.4. Review commercial company involvement in public sector data systems

Commercial company involvement in public sector data systems concerned participants in our focus groups and interviews as well as our survey. It also confused them, sometimes because of a lack of clear information about the nature of their involvement.

- **Recommendation:**

  8.4.1. Support or commission further research into the specific aspects of commercial company involvement in public sector data systems that are concerning and confusing, how to communicate complex public-private partnerships, and whether visualisations, of the kind we used and participants appreciated, can help to communicate complexity and overcome confusion.

9. References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. *ProPublica*. Retrieved from: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Bates J., Lin Y-W., Goodale P. (2016) 'Data journeys: Capturing the socio-material constitution of data objects and flows'. *Big Data & Society*. 10.1177/2053951716654502

Crenshaw, K. W. (1989) 'Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Color.' In: Martha Albertson Fineman, Rixanne Mykitiuk, Eds. *The Public Nature of Private Violence*. New York: Routledge.

D'Ignazio, C., & Bhargava, R. (2019). Data visualization literacy: A feminist starting point. *Data Visualization in Society*. Amsterdam: Amsterdam University Press.

Eubanks, V. (2017). *Automating inequality: How high-tech tools profile, police and punish the poor.* New York, NY: St Martins Press.

Fox, J., Pearce, K., Massanari, A., Riles, J., Szulc, Ł., Ranjit, Y., Trevisan, F., Soriano, C., Vitak, J., Arora, P., Ahn, S., Alper, M., Gambino, A., Gonzalez, C., Lynch, T., Williamson, L., Gonzales, A., (2021) 'Open Science, Closed Doors? Countering Marginalization through an Agenda for Ethical, Inclusive Research in Communication', *Journal of Communication*, 71 (5), pp 764–784, https://doi.org/10.1093/joc/jqab029

Kalluri, P. (2020) 'Don't ask if artificial intelligence is good or fair, ask if it shifts power', *Nature,* 7 July 2020, https://www.nature.com/articles/d41586-020-02003-2

Katwala, S. (2021) *Race and opportunity in Britain: finding common ground*, British Future, https://www.britishfuture.org/wp-content/uploads/2021/03/Race-and-opportunity-in-Britain.Final_.30.3.21.pdf

Kennedy, H. & Hill, R. (2017) 'The feeling of numbers: emotions in everyday engagements with data and their visualisation', *Sociology,* 52 (4): 830-848, http://journals.sagepub.com/doi/full/10.1177/0038038516674675.

Kennedy, H., Steedman, R., Jones, R. (2020) 'Approaching public perceptions of datafication through the lens of inequality: a case study in public service media' *Information, Communication and Society,* https://doi.org/10.1080/1369118X.2020.1736122.

Kennedy, H., Oman, S., Taylor, M., Bates, J., Steedman, R (2020b) Public understanding and perceptions of data practices: a review of existing research, https://livingwithdata.org/project/wp-content/uploads/2020/05/living-with-data-2020-review-of-existing-research.pdf

Kennedy, H., Taylor, M., Oman, S., Bates, J., Medina-Perea, I., Ditchfield, H. and Pinney, L. (2021) Context and prior awareness matter in public concerns about data uses: findings from the 2020 Living With Data survey, https://livingwithdata.org/project/wp-content/uploads/2021/07/Living-With-Data-2020-Survey-Summary-Report.pdf

Kennedy, H., Elgesem, D., & Miguel, C. (2015) On fairness: user perspectives on social media data mining. *Convergence*, 23 (3), 270-288, https://journals.sagepub.com/doi/abs/10.1177/1354856515592507

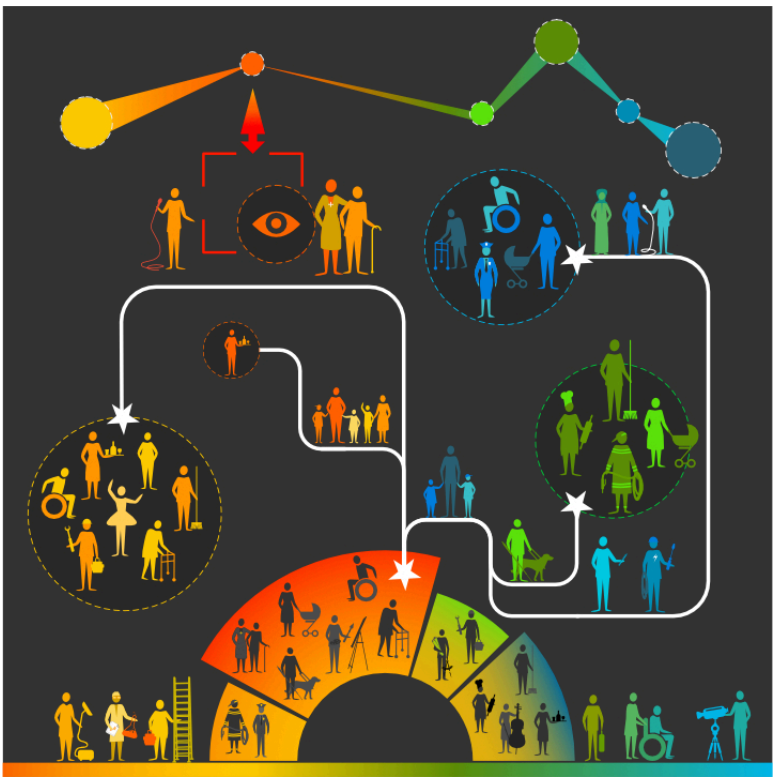
Law, J., & Urry, J. (2004) 'Enacting the social', *Economy and Society,* 33(3), 390–410.

Malik, S., Ryder, M., Marsden, S., Lawson, R. and Gee, M. (20210) *BAME:* A report on the use of the term and responses to it, BBC, https://bcuassets.blob.core.windows.net/docs/csu2021325-lhc-report--bbchighres231121-132836254614117870.pdf .

Medina Perea IA, Bates J & Cox A (2019) Using data journeys to inform research design: Socio-cultural dynamics of patient data flows in the UK healthcare sector. *iConference 2019 Proceedings*

Noble, S. (2018). *Algorithms of oppression: How search engines reinforce racism.* New York, NY: New York University Press.

Oman, S. (2021). *Understanding Well-being Data: Improving Social and Cultural Policy, Practice and Research*. Palgrave MacMillan.

Verma, S., and Rubin, J. (2018). Fairness Definitions Explained. In *FairWare'18: IEEE/ACM International Workshop on Software Fairness*, May 29, 2018, Gothenburg, Sweden

Living With Data is funded by The Nuffield Foundation. The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare, and Justice. It also funds student programmes that provide opportunities for young people to develop skills in quantitative and scientific methods. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics and the Ada Lovelace Institute. The Foundation has funded *Living With Data*, but the views expressed are those of the authors and not necessarily the Foundation.